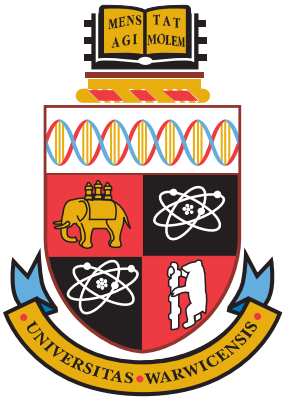
THE UNIVERSITY OF WARWICK



A Systems Biology Analysis of Feedback Control in Pheromone Signalling of Fission Yeast

by

Manuel Alejandro Esparza Franco

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy in Systems Biology

Systems Biology Doctoral Training Centre

February 2016

Declaration of Authorship

I, Manuel Alejandro Esparza Franco, declare that this thesis titled, 'A Systems Biology Analysis of Feedback Control in Pheromone Signalling of Fission Yeast' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

The following article was published as a result of the work contained in this thesis: Croft W, Hill C, McCann E, Bond M, Esparza-Franco MA, Bennet J, Rand D, Davey J and Ladds G. (2013). A physiologically required G protein-coupled receptor (GPCR)-Regulator of G protein signalling (RGS) interaction that compartmentalizes RGS activity. J. Biol. Chem., 288(38): 27327–27342.

Signed: _____

Date: _____

Abstract

Cell signalling comprises the systems used by cells to detect changes in their environment and to transduce the information into appropriate adjustments enforced by regulatory proteins. Due to its central role in all life processes, the study of cell signalling is a major focus of current biomedical research. The fission yeast *Schizosaccharomyces pombe* (*S. pombe*) is a single-celled organism used as a model to simplify the study of eukaryotic cell signalling, as it shares many features of interest with human cells. In this thesis a systems biology approach was used to investigate the roles of feedback regulation to control the dynamics of pheromone signalling in *S. pombe*. To this end, a quantitative dynamical model was built describing the pheromone-induced activation of the master transcription factor Ste11, as well as the coupled positive and negative feedback loops that arise from Ste11 activity. To constrain the model, a collection of data sets were generated by performing absolute quantification measurements of pheromone-dependent changes in the concentration of the model species. Structural identifiability analyses were used to select the measured species, while confidence intervals of the estimated parameters were determined through profile likelihood estimation. Analysis of the resulting model revealed a role for the pheromone signalling feedback loops to aid in the discrimination of different pheromone input doses. Through their combined action, feedback control defines the concentration and time thresholds in Ste11 activity that must be satisfied for the cell to commit to a sexual development fate.

Acknowledgements

I am deeply grateful and indebted to my supervisors Graham Ladds and Michael Chappell for their endless patience and support. I would like to acknowledge their invaluable contributions to this work.

I would like to thank all the people in the Ladds and Chappell groups that have helped me over the years.

To my wife, family, and friends.

Contents

List of Figures

List of Tables

Abbreviations

AIC	**A**kaike **I**nformation **C**riterion
BSA	**B**ovine **S**erum **A**lbumin
cAMP	**c**yclic **A**denosine **M**ono**P**hosphate
DMM	**D**efined **M**inimal **M**edia
DNA	**D**eoxyribo**N**ucleic **A**cid
dsDNA	**d**ouble-**s**tranded **DNA**
ddPCR	**d**igital **d**roplet **PCR**
ELISA	**E**nzyme-**L**inked **I**mmuno**S**orbent **A**ssay
FOA	5-**F**luoro**O**rotic **A**cid
GAP	**G**TPase **A**ccelerating **P**rotein
GDP	**G**uanosine **Di****P**hosphate
GEF	**G**uanosine nucleotide **E**xchange **F**actor
GFP	**G**reen **F**luorescent **P**rotein
GPCR	**G** **P**rotein-**C**oupled **R**eceptor
GTP	**G**uanosine **Tri****P**hosphate
HPLC	**H**igh **P**erformance **L**iquid **C**hromatography
LHS	**L**atin **H**ypercube **S**ampling
MAPK	**M**itogen **A**ctivated **P**rotein **K**inase
MAP2K	**MAPK K**inase
MAP3K	**MAP2K K**inase
mRNA	**m**essenger **RNA**
NMWL	**N**ominal **M**olecular **W**eight **L**imit
OD	**O**ptical **D**ensity
ODE	**O**rdinary **D**ifferential **E**quation

ONPG	**O**-**N**itro**P**henyl-D-**G**alactoside
ORF	**O**pen **R**eading **F**rame
PCR	**P**olymerase **C**hain **R**eaction
PL	**P**rofile **L**ikelihood
PLE	**P**rofile **L**ikelihood **E**stimation
qPCR	Real-time **q**uantitative **PCR**
RGS	**R**egulator of **G** Protein **S**ignalling
RNA	**R**ibo**N**ucleic **A**cid
RT-ddPCR	**R**everse **T**ranscription **ddPCR**
RT-qPCR	**R**everse **T**ranscription **qPCR**
SD	**S**tandard **D**eviation
STAUS	**S**imilarity **T**ransformation **A**pproach for **U**ncontrolled **S**ystems
S. pombe	*Schizosaccharomyces pombe*
UTR	**U**n**T**ranslated **R**egion
YE	**Y**east **E**xtract

Chapter 1

Introduction

1.1 Cell signalling

To remain adapted in a dynamic environment, cells must be able to perceive, interpret, and respond to external cues in a precise and timely manner. Extracellular signals trigger responses necessary for almost all cellular processes, including homeostasis, proliferation, differentiation, motility, and multicellular coordination (Gurdon *et al.*, 1993; Gilman, 1987; Falke *et al.*, 1997; Cargnello and Roux, 2011).

Signals are detected by receptor proteins, which become activated upon binding of the signal molecules. This in turn activates one or more signalling pathways, which transduce the signals inside the receiving cell and distribute them to appropriate intracellular targets (Alberts *et al.*, 2008).

Errors in signalling have been linked to many complex and prevalent human diseases, such as cancer, diabetes, and autoimmunity (Bazigou and Rallis, 2007; Ahrén, 2009), thus, understanding cell signalling is a major focus of current biomedical and pharmacological research (Lin and Liu, 2008).

1.1.1 Gene regulatory networks

Cell signalling systems convert variations in extracellular conditions into intracellular information that modifies the activity of key protein regulators. These regulators typically receive multiple and disparate inputs, and will exhibit very specific levels of activity based on the particular combination of inputs they receive, thus, essentially acting as logic processing systems.

However, higher level cellular decisions that result in qualitatively distinct behaviours, such as cell differentiation and commitment to a particular cell fate, require the coordinated action of multiple genes, functionally interconnected and capable of executing a specific cellular programme. These genetic circuits are known as gene regulatory networks (Davidson and Levine, 2005).

The specific arrangement in which genes are causally linked in a regulatory network defines its architecture, from which emergent properties arise, such as positive and negative feedback loops, that can only be completely understood at the system level (Karlebach and Shamir, 2008).

1.2 Models of eukaryote cell signalling

The study of cellular decisions directly in human or mammalian cells can become intractable because of the complexity caused by extensive feedback, crosstalk, and sheer size of the networks. Bacterial cells on the other hand bear little resemblance to eukaryote biochemistry. Thus, it is necessary that a model to study eukaryote cell signalling and gene regulatory networks be simple enough to allow studying properties of an individual pathway, with minimal interference from external signals, but remains relevant to a wide variety of organisms and signalling mechanisms.

Yeast are unicellular fungi that have been established as models of eukaryote cell biology due to their ease of culturing and genetic manipulation. In particular, the mating response pathway of the fission yeast *Schizosaccharomyces pombe* (*S. pombe*), provides an excellent opportunity to study regulatory mechanisms governing cell differentiation in relative isolation, thanks to the availability of mutant strains that can abolish many external influences, while combining several features that are of commercial, medical, and evolutionary interest (Davey, 1998).

1.3 Mating in *S. pombe*

S. pombe cells usually exist as haploids of either P (h^+) or M (h^-) mating types that proliferate through mitosis (Figure 1.1). However, upon nutritional depletion, cells of opposite mating types will fuse pairwise (conjugation), including fusion of the two nuclei (karyogamy), to produce diploid cells. The zygote undergoes meiosis and develops into an ascus containing four haploid ascospores. Spores are kept in a dormant state capable of tolerating severe environmental stresses, thus, sexual differentiation in yeast has evolved into a strategy to survive drastic environmental changes and hazardous stimuli. When suitable growth conditions are restored, the spores germinate to complete the life cycle (Leupold, 1950; Egel, 1971).

Sexual differentiation is energetically costly, and exiting the fast dividing haploid state when it is not absolutely necessary would be a great evolutionary disadvantage, therefore, the mating response genetic programme must be under tight control, to ensure it is only activated under appropriate situations. Many layers of regulation exist to control the commitment to sexual differentiation, all of which converge on the Ste11 transcription factor (Otsubo and Yamamoto, 2012).

Ste11 is a member of the high mobility group (HMG) transcription factor family that serves as a master switch to activate the mating response programme.

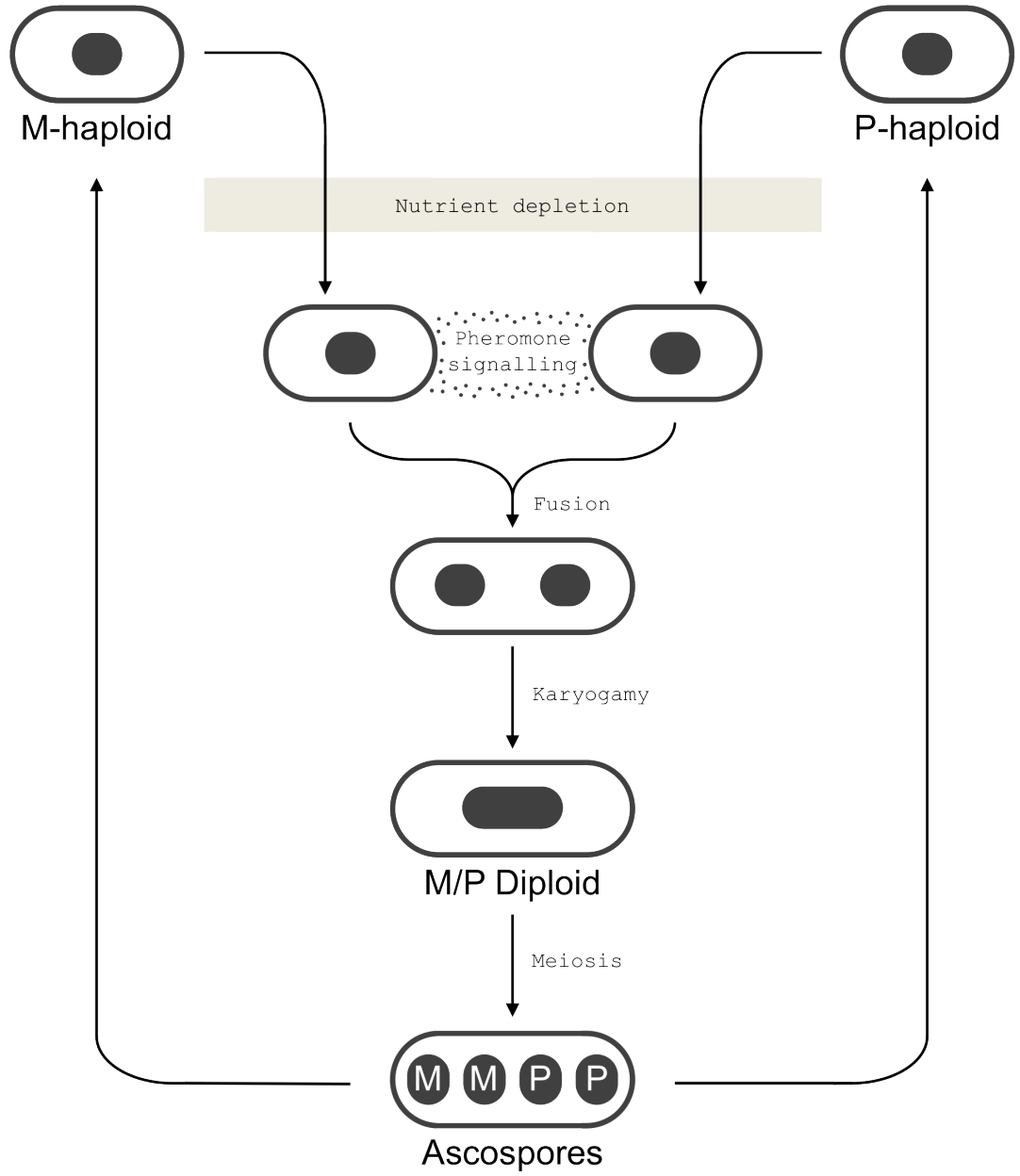


FIGURE 1.1: Life cycle of *S. pombe*.

Cells lacking the *ste11* gene are completely sterile, while overexpression of *ste11* induces sexual differentiation regardless of environmental conditions (Sugimoto *et al.*, 1991). Ste11 controls the expression of at least 78 genes that are required to produce the physiological and morphological changes necessary to mate (Mata and Bähler, 2006; Xue-Franzén *et al.*, 2006). The regulation of Ste11 expression and activity is complex, including suppression during mitotic growth by nutrient-rich conditions, cell cycle phase, nuclear exclusion, protein instability, and translation

inhibition of itself and its targets, as well as activation by stress and pheromone communication (Anandhakumar *et al.*, 2013).

1.3.1 Nutrient-sensing

Sexual development in fission yeast culminates in the formation of dormant ascospores which can survive through adverse conditions. Thus it makes sense for nutrient-sensing mechanisms to have a direct input into the control of Ste11, to ensure that mitotic growth is maintained when sufficient nutrients are available, and that mating and sporulation occurs only when resources become limiting. Two distinct nutrient-sensing pathways are known to regulate *ste11* expression, the cyclic adenosine monophosphate (cAMP) pathway, and the target of rapamycin (TOR) complex 1 (TORC1) pathway.

1.3.2 cAMP pathway

S. pombe cells monitor the availability of extracellular carbon through the putative glucose receptor Git3 (Hoffman, 2005). In the presence of sufficient nutrients, the Git3 GPCR will activate its cognate heterotrimeric G protein, allowing the α-subunit Gpa2 to directly activate adenylate cyclase Cyr1 (Ivey and Hoffman, 2005), which catalyses the production of cAMP from ATP (Figure 1.2) (Kawamukai *et al.*, 1991). Analysis of cAMP-related mutants has revealed that cells with high amounts of intracellular cAMP are deficient for mating and sporulation (DeVoti *et al.*, 1991; Mochizuki and Yamamoto, 1992), whereas cells defective in cAMP production will initiate sexual development even under nutrient-rich conditions (Maeda *et al.*, 1990; Kawamukai *et al.*, 1991).

cAMP controls mating initiation by modulating the activity of the cAMP-dependent protein kinase (PKA) through its regulatory subunit Cgs1 (Figure 1.2) (DeVoti

et al., 1991; Maeda *et al.*, 1994). In turn, PKA negatively regulates the transcription factor Rst2 by phosphorylation and nuclear exclusion (Kunitomo *et al.*, 2000; Higuchi *et al.*, 2002). Rst2 is a zinc-finger protein that induces the expression of *ste11* by binding to a STREP motif in the *ste11* promoter (Kunitomo *et al.*, 2000). Cells lacking *rst2* can grow mitotically but are sterile; this deficiency can be rescued by artificial expression of *ste11* (Kunitomo *et al.*, 2000).

Thus, exhaustion of carbon sources leads to the inactivation of Cyr1 and a subsequent depletion of intracellular cAMP, which liberates Rst2 from PKA regulation and promotes the transcription of *ste11* (Figure 1.2).

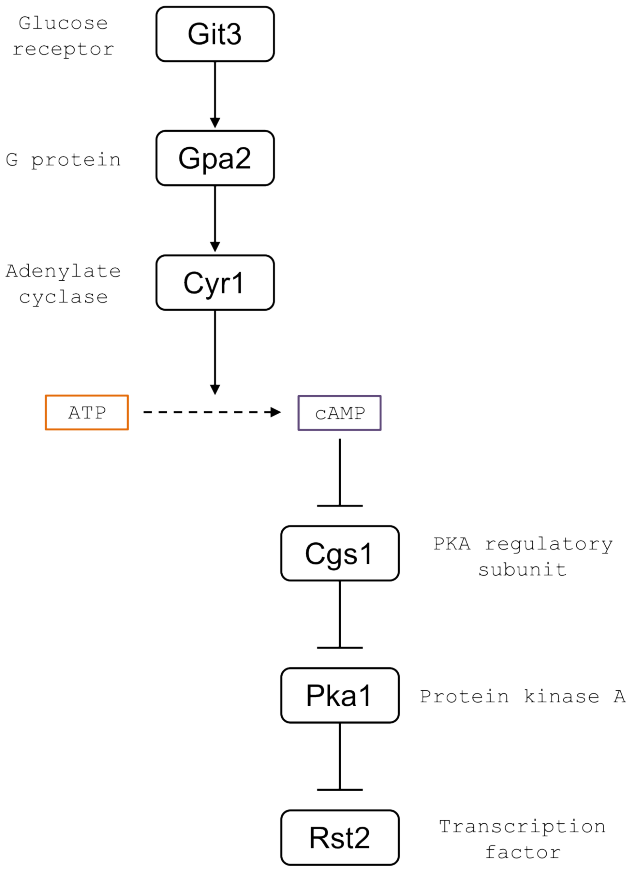


FIGURE 1.2: Nutritional regulation of *ste11* transcription. Rst2-mediated transcription of *ste11* is coupled to the availability of carbon through the cAMP pathway. cAMP production depends on the activation of adenylate cyclase (Cyr1) by the carbon sensor Git3-Gpa2. Intracellular levels of cAMP then modulate the activity of Pka1 (via Cgs1) to prevent *ste11* transcription when carbon sources are abundant.

1.3.3 TORC1 pathway

TORC1 is a multisubunit complex that contains the Tor2 kinase, a homologue of mammalian TOR (Weisman and Choder, 2001). The TORC1 pathway cooperates with PKA to repress Ste11 expression and activity; while the cAMP pathway surveys the availability of carbon sources, the TORC1 pathway is regulated by nitrogen sources (Uritani *et al.*, 2006).

Although the molecular details of how this pathway is activated, and how it regulates Ste11 activity have not been completely elucidated, overexpression of *tor2* will abolish the induction of *ste11* mRNA in response to starvation rendering cells sterile; whereas loss of *tor2*, or inactivation of Tor2 temperature-sensitive mutants will mimic nitrogen starvation, and will upregulate the expression of *ste11* and its downstream targets, causing the entry of cells into mating behaviour and meiosis even if grown in rich medium (Alvarez and Moreno, 2006; Uritani *et al.*, 2006; Weisman *et al.*, 2007; Matsuo *et al.*, 2007; Valbuena and Moreno, 2010).

These results suggest that Tor2 and PKA kinases play overlapping roles to repress Ste11 activity and sexual development. However, even Tor2 overexpression is not able to prevent *ste11* transcription if PKA becomes inactive, nor can it prevent nuclear accumulation of Ste11 in this situation (Valbuena and Moreno, 2010), showing that the cAMP pathway plays a more dominant role in regulating Ste11 activity.

1.3.4 Stress-responsive MAPK pathway

As mentioned above, ascospores are highly resistant to a variety of adverse conditions in addition to starvation. Therefore, other kinds of stresses are also able to promote entry into meiosis through regulation of Ste11. Fission yeast possess a core environmental stress response (CESR) which is deployed in response to any

kind of externally induced stress (Chen *et al.*, 2003), including heat, osmolarity, nutritional, and oxidative stress (Millar *et al.*, 1995; Shiozaki and Russell, 1995; Degols *et al.*, 1996; Shiozaki and Russell, 1996).

Detection of stress cues will converge in the response regulator Mcs4, making it bind to the Wis4 and Win1 MAP3Ks (Shiozaki *et al.*, 1997; Shieh *et al.*, 1997; Morigasaki *et al.*, 2013), creating a signalling scaffold that phosphorylates the MAP2K Wis1, which in turn phosphorylates the MAPK Sty1, a homologue of mammalian MAPK p38 (Millar *et al.*, 1995; Shiozaki and Russell, 1995; Samejima *et al.*, 1997; Shieh *et al.*, 1998; Samejima *et al.*, 1998). Activated Sty1 accumulates in the nucleus and phosphorylates several targets, including the bZIP transcription factor Atf1 (Shiozaki and Russell, 1996; Wilkinson *et al.*, 1996; Gaits *et al.*, 1998), this phosphorylation allows Atf1 to create a heterodimer with another bZIP protein, Pcr1 (Watanabe and Yamamoto, 1996; Kanoh *et al.*, 1996). The Atf1-Pcr1 dimer binds to the promoters of stress-response genes and also recruits Sty1 to those sites (Reiter *et al.*, 2008; Sansó *et al.*, 2011). This recruitment is key for the regulation of *ste11* expression by the stress-responsive pathway as explained below.

The activity of RNA polymerase II (PolII) can be controlled through the phosphorylation pattern of its C-terminal domain (CTD) (Phatnani and Greenleaf, 2006), and more specifically, it has been shown that phosphorylation of Ser-2 of PolII CTD is necessary for a robust induction of *ste11* expression (Coudreuse *et al.*, 2010; Sukegawa *et al.*, 2011). This phosphorylation is controlled by the CTDK-I kinase Lsk1, which is itself a target of Sty1. Thus, it is thought that when Sty1 is recruited to the *ste11* promoter via Atf1-Pcr1, it is able from this position to recruit and activate Lsk1 to promote phosphorylation of Ser-2 of PolII CTD during transcription initiation, which enhances the occupancy of PolII at the *ste11* locus (Coudreuse *et al.*, 2010; Sukegawa *et al.*, 2011).

1.3.5 Protein stability

Molecules with key regulatory roles typically have a high turnover rate, to allow a rapid control over their concentration (Alberts *et al.*, 2008). Ste11 is an unstable protein with a half-life of less than 20 min (Kjærulff *et al.*, 2007). Ste11 degradation is mediated by polyubiquitination and subsequent 26S proteasome destruction (Kitamura *et al.*, 2001; Kjærulff *et al.*, 2007). Importantly, this mechanism operates independently of phosphorylation status of Ste11 (Kjærulff *et al.*, 2007).

1.3.6 Cell cycle control

Sexual differentiation in yeast, and eukaryote cell differentiation in general, occurs only during the G_1 phase of the cell cycle, before the cell has committed itself to another round of division. Accordingly, transcription of Ste11 targets has been shown to be confined to G_1 (Stern and Nurse, 1998). This cell cycle constraint is enforced by the cyclin dependent kinase Cdc2, a homologue of mammalian Cdk1 (Lee and Nurse, 1987).

Cdc2 phosphorylates Ste11 at Thr82, which is located within the HMG-box DNA-binding motif of Ste11 (Sugimoto *et al.*, 1991; Kjærulff *et al.*, 2007). This phosphorylation impairs the ability of Ste11 to bind DNA by lowering its affinity to its target sequence by two to three fold (Kjærulff *et al.*, 2007), thus reducing its effectiveness to act as a transcription factor.

Since Cdc2 is cell-cycle regulated, having a minimum of activity during early G_1 (Stern and Nurse, 1996), this mechanism helps to constrain Ste11-dependent transactivation to the G_1 phase, where newly synthesized Ste11 can escape Cdc2 downregulation. However, this mechanism cannot fully prevent Ste11 activity, as cells carrying a Ste11 mutation that mimics constant phosphorylation at Thr82

(Ste11^{T82D}) are only semi-sterile, and Ste11^{T82D} is still able to weakly induce the expression of target genes (Kjærulff *et al.*, 2007).

1.3.7 Subcellular localization

Ste11 contains a basic cluster nuclear localization signal (NLS) in the HMG-box domain (Qin *et al.*, 2003), and so newly synthesized Ste11 will be shuttled into the nucleus. However, in growing cells, where Ste11 is produced at very low levels, its localization appears to be pancellular, and is not confined to the nucleus (Qin *et al.*, 2003). Treatment with leptomycin B, an exportin inhibitor, will cause nuclear accumulation of Ste11 in these cells (Qin *et al.*, 2003), indicating that Ste11 is also subject to active export from the nucleus.

Nuclear accumulation of Ste11 is required to trigger the sexual developmental switch, but this requires both starvation and pheromone stimulation (Qin *et al.*, 2003). In order to prevent nuclear accumulation of Ste11 in the absence of appropriate signals, the Pat1 kinase, which resides in the nucleus (McLeod *et al.*, 2000; Matsuyama *et al.*, 2006), phosphorylates Ste11 at positions Thr173 and Ser218 (Li and McLeod, 1996). This phosphorylation pattern allows Rad24, a 14-3-3 homologue (Ford *et al.*, 1994), to bind Ste11 and restrict its localization to the cytoplasm for the remainder of its lifetime (Kitamura *et al.*, 2001).

1.3.8 Pheromone response MAPK pathway

Upon early starvation, especially of nitrogen, Ste11 directs the transcription of mating-type-specific genes, including mating pheromones and pheromone receptors (Davey, 1998). Cells of mating type P produce P-factor pheromone, an unmodified 23 amino acid peptide, while mating type M cells produce M-factor pheromone, a 9 amino acid peptide that is farnesylated and carboxymethylated at

the C-terminus (Imai and Yamamoto, 1994; Davey, 1992). Mating pheromones are secreted into the extracellular space and will only bind to specific receptors present in cells of opposite mating type. Yeast cells are not motile, so they reach mating partners by growing elongated conjugation tubes (shmoo) towards the nearest source of appropriate pheromone (Fukui *et al.*, 1986; Davey, 1991).

M-cells detect P-factor through the Mam2 receptor, and P-cells detect M-factor through the Map3 receptor, both of which are G protein-coupled receptors (GPCRs) that dock to the same Gα protein Gpa1 (Figure 1.3) (Kitamura and Shimoda, 1991; Tanaka *et al.*, 1993; Obara *et al.*, 1991).

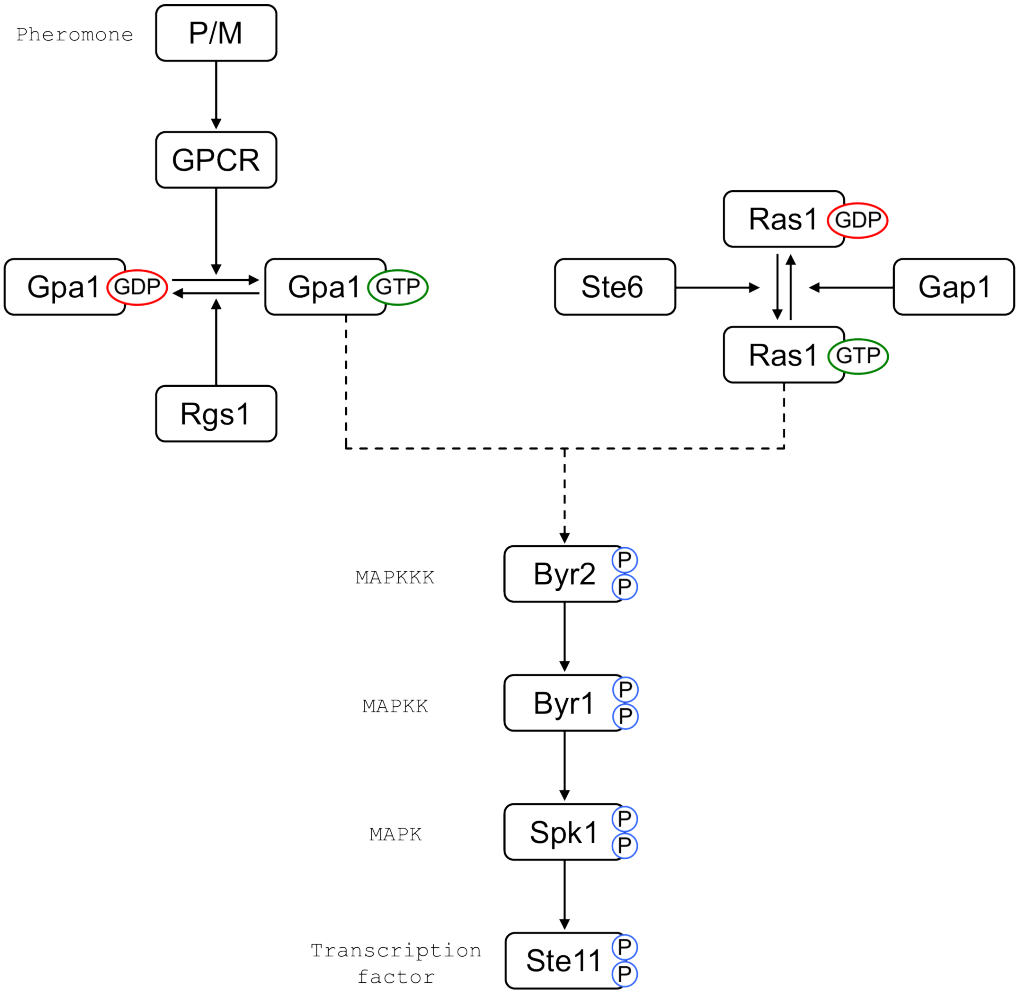


FIGURE 1.3: Pheromone response pathway of *S. pombe*. Mating-type specific receptors detect nearby compatible partners by pheromone concentration gradients. Receptors couple pheromone binding to G protein activation leading to the activation of a MAP kinase cascade, which begins the transcriptional reprogramming necessary for mating.

Pheromone binding to the receptor leads to the activation of the Gα, Gpa1 (Obara *et al.*, 1991), then, with the help of Ras1 (Xu *et al.*, 1994), activated Gpa1 will direct the activation of a transcriptional response, and a morphological response (Papadaki *et al.*, 2002). Signalling to initiate the transcriptional response is transduced through a MAP kinase cascade comprising the sequential phosphorylation of Byr2 (MAP3K), Byr1 (MAP2K), and Spk1 (MAPK) (Neiman *et al.*, 1993). Spk1 is then able to phosphorylate Ste11 at positions Thr305 and Thr317 (Kjærulff *et al.*, 2005). Although the precise consequence of these phosphorylations has not yet been determined, phosphomimetic mutations at these residues produce a hyperactive Ste11 allele that accelerates the speed of the mating response (Kjærulff *et al.*, 2005), presumably by bypassing its cell cycle or subcellular localization control.

1.3.9 Translational control

A final layer of regulation is thought to control Ste11 activity at the translational level. Nrd1 is an RNA-binding protein that contains 4 highly conserved RNA recognition motifs (RRMs) and has a preferential binding affinity for poly(U) tracts (Tsukahara *et al.*, 1998). Nrd1 represses the translation of Ste11-regulated genes until starvation has occurred, but it is unclear how specificity is achieved or how the actual mechanism of repression works. Genetic analyses of mutants to assess epistatic relationships have shown that Nrd1 acts independently of nutrient sensing and stress responsive pathways, however, activated Spk1 phosphorylates Nrd1 to repress its function (Tsukahara *et al.*, 1998; Oowatari *et al.*, 2011), thus placing this regulatory mechanism directly under pheromone signalling control.

1.4 Systems biology and the modelling loop

Systems biology is an approach to biological research that focuses on the integration of knowledge and data into mathematical models, to allow a systems level investigation of emergent behaviours that cannot be understood by studying the individual components of the system in isolation (Le Novère, 2015).

Mathematical models are a description of an experimentally delineated phenomenon by means of mathematics, with a view to capture the salient aspects of the phenomenon at hand (van den Berg, 2011). Although not immediately clear from this definition, mathematical models have proved to be a powerful tool to help integrate information from large and varied data sets, gain mechanistic understanding of a system, and predict the system response in a wide range of alternative scenarios (Chandran *et al.*, 2008; Zheng and Sriram, 2010; MacDonald *et al.*, 2011).

The goals of a mathematical model can be varied, and the usefulness of a particular model must be judged on the basis of those objectives. Models can be used to estimate the numerical value of parameters of interest, such as the rate of an enzymatic reaction. Alternatively, models can make predictions of data values for scenarios that have not been tested, or molecular species that cannot be easily measured. A third option is to construct several models that implement competing hypotheses to describe the same phenomenon, and then perform experiments that can discriminate between them. Although the options listed are not mutually exclusive, the emphasis for a particular model will usually fall on only one of these categories, however, they can all be seen from a wider perspective as being maps used to locate areas of ignorance.

Building an appropriate mathematical model for a system of interest relies on prior knowledge, model objectives, and available data (Ljung, 1987). Model building

is never a linear process, and typically requires several iterations between experimentation and model refinement to reach the model objectives.

1.4.1 Model and parameter identifiability

Mathematical modelling can be used to predict the behaviour of a signalling pathway and the effects of altering its structure; however, the strength of its predictions depends intrinsically on the accuracy of estimated model parameters. Due to technical limitations, biological reaction networks are often only partially observable. This means that not all species incorporated into a model can be measured directly. Given a certain amount and quality of experimental data, the estimation of the model parameters might be ambiguous, or non-identifiable. If model parameters are not well determined, it follows that model predictions will not be either (Kreutz and Timmer, 2009).

Two types of non-identifiability can occur. A structural non-identifiability arises from the model structure only, and is independent of any experimental data (Bellman and Åström, 1970). A parameter that is structurally identifiable may still be practically non-identifiable, owing to insufficient amount and quality of experimental data or inappropriately chosen measurement time points (Raue *et al.*, 2011).

The functional relationships that cause structural non-identifiabilites are a consequence of the model structure only, due to insufficient mapping between observables and state variables, and thus can only be solved by factors affecting this mapping, such as qualitatively different measurements, or model reduction.

Various approaches have been proposed to detect non-identifiability. Approaches that analyse model equations analytically are called a priori methods, as they allow the testing of identifiability before experimental data are available. These

methods have the advantage that conclusions about identifiability hold globally for the entire parameter space (Chis *et al.*, 2011); however, these approaches only test for structural identifiability. Although structural identifiability might be ensured a priori, practical non-identifiability can cause severe problems when estimating model parameters from real experimental data (Raue *et al.*, 2011).

Data-based approaches utilise the shape of the likelihood function to infer identifiability. This naturally involves experimental data, and therefore allows statements about practical identifiability akin to confidence intervals (Hengl *et al.*, 2007). Owing to data dependency, the results of these methods cannot ensure global validity. Nevertheless, they allow statements about the region of parameter space specified by experimental data, which is often sufficient for applications (Raue *et al.*, 2009).

If non-identifiability is detected, it can be resolved either by experimental design, measuring additional data under suitable conditions; or by model reduction, tailoring the size of the model to the information content provided by the experimental data (Roper *et al.*, 2010).

1.5 Project aims

As outlined in section 1.3, the regulation of the master transcription factor Ste11 is multilayered and complex, therefore, to date there have been no attempts to model the dynamics of the mating response in *S. pombe*. This lack of mathematical modelling efforts has prevented the study of system-level properties that may control the decision making process for commitment to mating.

To address this issue, the use of mutant strains that cut down the complexity of Ste11 regulation is proposed, which would allow the mating response to be modelled exclusively as a function of the pheromone signalling pathway and its known feedback loops. The rationale is that by creating a model of a simplified

mating response, the system-level mechanisms that govern the decision to mate can begin to be studied, while minimising the uncertainty introduced from the lack of mechanistic knowledge about all the pathways involved in Ste11 regulation. In addition, this new model could serve as a starting point for future models that re-incorporate the control mechanisms ignored in this study.

To accomplish these goals, the following project aims were established:

- To develop a quantitative dynamical model of *S. pombe* pheromone signalling that takes into account feedback loops at the gene regulatory network level.

- To generate the necessary quantitative time course data to estimate the parameters of the model.

- To analyse the model to investigate the roles of feedback regulation in controlling the dynamics of pheromone signalling.

Chapter 2

Materials and methods

2.1 Experimental methods

2.1.1 Bacterial transformation

Escherichia coli strain DH5α (*supE44 hsdR17 endA96 thi-1 relA1 recA1 gyrA96*) was used to prepare chemically competent bacteria and for DNA transformation as described by Sambrook and Russell (2001).

2.1.2 Small scale isolation of plasmid DNA

Plasmid DNA was isolated and purified from transformed bacterial cultures using QIAprep Spin Miniprep kits according to manufacturer instructions (Qiagen, UK).

2.1.3 Amplification of DNA sequences by polymerase chain reaction (PCR)

PCR reactions were performed using a Phusion High-Fidelity DNA polymerase (ThermoFisher, UK). Oligonucleotide primers were prepared by standard solid-phase synthesis, and purified by desalting (Sigma-Aldrich, UK). Reactions were setup in ice and transferred to a pre-heated thermocycler for the amplification reaction. Typical reactions were prepared as detailed in Table 2.1. Typical cycling conditions are shown in Table 2.2

Table 2.1: Phusion PCR reaction setup.

Component	Final concentration
5X Phusion HF Buffer	1X
dNTPs	200 µM
Forward primer	0.5 µM
Reverse primer	0.5 µM
Template DNA	$\sim$ 10 ng
Phusion DNA Polymerase	1 unit / 50 µL
Nuclease-free water	to 50 µL

2.1.4 Site-directed mutagenesis

Site-directed mutagenesis of DNA sequences was performed by PCR as described by Liu and Naismith (2008). The main difference in this protocol compared to other methods is the primer design stage, where the primers should have an overlapping region at their 5' end, and a large non-overlapping region at their 3' end. The difference in melting temperature between the two regions should be 5 to 10°C higher for the non-overlapping sequences. The protocol then proceeds

Table 2.2: Thermal cycling conditions for Phusion PCR. The optimal annealing temperature was determined empirically for each primer pair.

Step	Temperature	Time
Initial denaturation	98°C	30 sec
	98°C	10 sec
25-35 cycles	X°C	30 sec
	72°C	30 sec per kb
Final extension	72°C	5-10 min
Hold	4	

through PCR amplification of the target plasmid with the mutagenesis primers, followed by DpnI digestion to eliminate the original template DNA, and bacterial transformation with the mutated PCR product.

2.1.5 Gibson assembly

Gibson assembly allows the joining of DNA molecules with overlapping sequences in a single isothermal reaction by simultaneous use of exonuclease, polymerase, and ligase enzymes. Molecular cloning through Gibson assembly was performed using the Gibson Assembly Master Mix (New England BioLabs, UK). DNA fragments for Gibson assembly were prepared by PCR as described above. Reactions were prepared with 100 ng of vector and a 2-3 fold concentration excess of inserts in a 20 µL volume with 1X final concentration of the Master Mix. The assembly reactions were performed in a thermocycler at 50°C for 30 min, and stored at 4°C until used for bacterial transformation.

2.1.6 DNA sequencing

All DNA constructs created in this work were verified by Sanger sequencing (GATC Biotech, UK).

2.1.7 RNA extraction from *S. pombe*

The protocol used for RNA extraction from fission yeast was first described by Lyne *et al.* (2003). Cells were harvested by centrifugation at 800 × *g* for 2 min, and snap frozen in dry ice/methanol. Frozen pellets were thawed on ice, resuspended in 1 mL of RNase-free water (Fisher Scientific, UK), and transferred into 2 mL PCR clean safe-lock tubes (Eppendorf, UK). Cells were centrifuged at 1,500 × *g* for 10 sec and the supernatant removed. Cell pellets were resuspended in 750 µL TES (10 mM Tris pH 7.5; 10mM EDTA pH 8; 0.5% SDS), followed immediately by adding 750 µL of acidic phenol-chloroform (Sigma-Aldrich, UK), vortex mixing, and incubation at 65°C for 1 hr. After heat treatment, samples were cooled on ice for 1 min, mixed by vortex for 20 sec, and centrifuged at 15,000 × *g* for 15 min at 4°C. Heavy phase lock gel 2 mL tubes (5 Prime, Germany) were prepared by centrifugation for 10 sec at max rpm and loaded with 700 µL of acidic phenol-chloroform. 700 µL of the water phase from sample tubes were transferred to the phase lock gel tubes, mixed thoroughly but gently through repeated inversion, and phases were separated by centrifugation at 15,000 × *g* for 5 min at 4°C. 700 µL of the water phase were transferred to a new batch of phase lock gel tubes, this time prepared with 700 µL of chloroform:isoamyl alcohol (24:1) (Sigma-Aldrich, UK). Samples were again mixed gently, and phases were separated by centrifugation at 15,000 × *g* for 5 min at 4°C. 500 µL of the water phase were transferred to normal 2 mL tubes containing a mixture of 1.5 mL 100% ethanol pre-chilled to -20°C and 50 µL of 3 M sodium acetate pH 5.2. RNA was allowed to precipitate overnight at -20°C. Samples were then centrifuged at room temperature for 10 min at 15,000

× *g*, washed with 500 µL of 70% ethanol pre-chilled to 4°C, and air dried at room temperature for 5 min. RNA pellets were resuspended in 100 µL of RNase-free water by allowing it to soak at room temperature for 10 min, followed by dissolving through pipetting and gentle vortex mixing for 10 sec.

Extraction yields were quantified by spectrophotometry with a NanoDrop 2000c instrument (Thermo Scientific, UK), and 100 µg of total RNA were further purified using a RNeasy plus kit (Qiagen, UK). Briefly, genomic DNA is first removed by filtering the RNA solution through a proprietary column that selectively binds DNA. The flowthrough is then mixed with 1 volume of 70% ethanol and bound to a silica gel spin column by centrifugation at 8,000 × *g*. Bound RNA is washed twice with 500 µL of buffer RPE, and eluted twice in 50 µL of RNase-free water. Total RNA integrity was verified with an Agilent 2100 Bioanalyzer nanofluidics device running Agilent RNA 6000 Nano chips (Agilent Technologies, UK).

2.1.8 cDNA synthesis

Reverse transcription (RT) from total RNA was performed using the QuantiTect Reverse Transcription kit (Qiagen, UK), following manufacturer recommendations. Up to 1 µg of total RNA was used per reaction and prior to the reverse transcription reaction, genomic DNA was eliminated by incubation with the proprietary gDNA Wipeout buffer at 42°C for 2 min. The entire gDNA elimination reaction was then used for cDNA synthesis with an RT primer mix consisting of a blend of oligo-dT and random primers.

2.1.9 Real-time quantitative PCR (qPCR)

qPCR assays were prepared using the QuantiTect SYBR Green PCR kit (Qiagen, UK), following manufacturer recommendations. Reactions were setup in 25 µL

volumes with a final primer concentration of 0.3 µM for each one, and 100 ng of cDNA per reaction. Data acquisition was performed on a ABI PRISM 7000 Sequence Detection System (Applied Biosystems, UK), using semi-skirted 96-well plates (Starlab, UK), sealed with ThermalSeal RT2 adhesive films (Alpha Labs, UK). The cycles of quantification (C_q) were determined through the SDS software that controls the thermocycler.

2.1.10 Droplet digital PCR (ddPCR)

ddPCR assays were prepared using the QX200 ddPCR EvaGreen Supermix kit (Bio-Rad, UK), following manufacturer recommendations. Reactions were setup in 20 µL volumes with a final primer concentration of 0.3 µM for each one, and a 1:20 dilution of cDNA sample per reaction, corresponding to $\sim 40 - 50$ ng of cDNA. Assays were run using the QX200 ddPCR system (Bio-Rad, UK), which consists of the QX200 droplet generator, the QX200 droplet reader, and the QuantaSoft software for data analysis. PCR amplification was performed on a C1000 Touch thermal cycler (Bio-Rad), using semi-skirted 96-well plates (Starlab, UK), sealed with a PX1 PCR plate sealer (Bio-Rad, UK).

2.1.11 Concentrating protein samples

The concentration of protein samples was performed using centrifugal filters for ultrafiltration. Samples were concentrated by a factor of 10X using the 500 µL Amicon Ultra-0.5 filters (Millipore, UK), or by a factor of 200X using the 70 mL Centricon Plus-70 filters (Millipore, UK). The nominal molecular weight limit (NMWL) of the filters was chosen to be at least two times smaller than the molecular weight of the protein intended for concentration. For Sxa2-GFP samples (∼50 kDa) a 10,000 NMWL was used. Samples were centrifuged for 30 min at max.

speed, followed by a recovery step inverting the filter into a collection tube, and centrifuging at 1,000 *g* for 1 min.

2.1.12 anti-GFP Enzyme-linked immunosorbent assay (ELISA)

Quantification of GFP fusion proteins was performed using the ab171581 - GFP SimpleStep ELISA kit according to manufacturer instructions (abcam, UK). Readings of optical density (OD) were performed at 420 nm using a Berthold Mithras LB940 microplate reader (Berthold Technologies, UK).

2.1.13 Yeast strains

The yeast strains used in this study are listed in Table 2.3. Correct integration of DNA fragments used to generate new yeast strains was verified by PCR analysis.

2.1.14 Yeast growth conditions

All strains used in this study were cultured at 29°C. Liquid cultures were additionally aerated by constant shaking at 180 rpm using New Brunswick Innova 44 shaking incubators (Eppendorf, UK). Yeast growth media was prepared following Davey *et al.* (1995). For routine growth and maintenance yeast were cultured using yeast extract (YE) rich medium (0.5% yeast extract, 3% glucose). For pheromone experiments yeast were cultured using defined minimal media (DMM) (Table 2.4 and 2.5). Required supplements for auxotrophies were added to a final concentration of 225 mg/L as required. To prepare solid growth media, liquid media was supplemented with 2% agar (Merck BDH, UK). P-factor used for pheromone experiments was custom-ordered and was synthesised using 9-fluoroenylmethoxycarbonyl (FMOC) protection chemistry (Alta Bioscience Ltd, UK). The final concentration of P-factor was varied, and is explicitly indicated for each experiment in the results.

TABLE 2.3: *S. pombe* strains used in this study.

Strain	Genotype	Reference
JY486	*mat1-M*, Δ*mat2/3::LEU2*$^-$, *leu1*$^-$, *ade6-M216*, *ura4-D18*, *cyr1-D51*	(Didmon *et al.*, 2002)
JY522	*mat1-M*, Δ*mat2/3::LEU2*$^-$, *leu1*$^-$, *ade6-M216*, *ura4-D18*, *cyr1-D51*, *sxa2::ura4+*	(Didmon *et al.*, 2002)
JY544	*mat1-M*, Δ*mat2/3::LEU2*$^-$, *leu1*$^-$, *ade6-M216*, *ura4-D18*, *cyr1-D51*, *sxa2>lacZ*	(Didmon *et al.*, 2002)
JY1325	*mat1-M*, Δ*mat2/3::LEU2*$^-$, *leu1*$^-$, *ade6-M216*, *ura4-D18*, *cyr1-D51*, *sxa2>GFP*	(Smith, 2009)
JY1741	*mat1-M*, Δ*mat2/3::LEU2*$^-$, *leu1*$^-$, *ade6-M216*, *ura4-D18*, *cyr1-D51*, *sxa2*^{S200A}*-6xHis*	This study
JY1743	*mat1-M*, Δ*mat2/3::LEU2*$^-$, *leu1*$^-$, *ade6-M216*, *ura4-D18*, *cyr1-D51*, *sxa2-GFP*	This study
JY1744	*mat1-M*, Δ*mat2/3::LEU2*$^-$, *leu1*$^-$, *ade6-M216*, *ura4-D18*, *cyr1-D51*, *sxa2*^{S200A}*-GFP*	This study

2.1.15 Yeast transformation

S. pombe cells were transformed using the lithium acetate method as described by Okazaki *et al.* (1990).

2.1.16 Cell counting and cell volume measurement

The number of cells per mL of culture and the median volume of cells was determined using a Z2 Coulter particle counter (Beckman Coulter, UK).

Table 2.4: Composition of DMM growth medium.

Per litre	Final concentration
5 g NH_4Cl	93.5 mM
2.2 g Na_2HPO_4	15.5 mM
3 g potassium hydrogen phthalate	14.7 mM
20 g glucose	2% w/v
20 mL salt stock (see Table 2.5)	
1 mL vitamin stock (see Table 2.5)	
0.1 mL mineral stock (see Table 2.5)	

2.1.17 β-galactosidase assay

Ste11-mediated transcriptional activity in response to P-factor was measured using strain JY544 containing a *lacZ* reporter integrated at the *sxa2* locus as described by Didmon *et al.* (2002). Cells were stimulated with the desired dose of P-factor and incubated under standard conditions for 16 hrs. β-galactosidase activity was determined by lysing 50 µL of cells in 750 µL of Z-buffer (Table 2.6), and continuing the incubation at 29°C for 90 min. Reactions were stopped with 200 µL of 2 M sodium carbonate. OD measurements were read at 420 nm and activity units were expressed as $OD_{420}/10^6$ cells.

2.1.18 Confocal fluorescence microscopy

Cells were imaged on DMM agarose pads (2 %) using a CoverWell imaging chamber (Grace Bio-Labs, USA). Coverslips were sealed using vaseline to prevent drying. Image acquisition was performed using a Leica TCS SP5 confocal microscope (Leica Microsystems, UK). Microscopy images were processed using the open source FIJI implementation of ImageJ (Schindelin *et al.*, 2012).

Table 2.5: Stock solutions for DMM media.

50× salt stock (per litre)	Final concentration
52.5 g $MgCl_2 \cdot 6H_2O$	0.26 M
0.735 g $CaCl_2 \cdot 2H_2O$	4.99 mM
50 g KCl	0.67 M
2 g Na_2SO_4	14.1 mM
1000× vitamin stock (per litre)	
1 g pantothenic acid	4.20 mM
10 g nicotinic acid	81.2 mM
10 g inositol	55.5 mM
10 mg biotin	40.8 µM
10,000× mineral stock (per litre)	
5 g boric acid	80.9 mM
4 g $MnSO_4$	23.7 mM
4 g $ZnSO_4 \cdot 7H_2O$	13.9 mM
2 g $FeCl_2 \cdot 6H_2O$	7.4 mM
0.4 g molybdic acid	2.47 mM
1 g KI	6.02 mM
0.4 g $CuSO_4 \cdot 5H_2O$	1.60 mM
10 g citric acid	47.6 mM

2.2 Computational methods

Model building, simulations and sensitivity analysis were performed using the SimBiology toolbox for MATLAB R2013a (The MathWorks, UK). Parameter estimation and Profile Likelihood estimation was performed using the Potterswheel (Maiwald and Timmer, 2008), and Data2Dynamics (Raue *et al.*, 2015) toolboxes

Table 2.6: Composition of Z-buffer for β-galactosidase assays.

Component	Final concentration
$NaPO_4$	0.1 M
KCl	10 mM
$MgSO_4$	1 mM
β-mercaptoethanol	50 mM
chloroform (v/v)	0.5%
SDS (w/v)	0.005%

for MATLAB. Symbolic calculations necessary for structural identifiability and indistinguishability analysis were performed using Maple 2015 (Maplesoft, UK).

Chapter 3

Model derivation

3.1 Existing models of pheromone signalling

To date the only published examples of using mathematical modelling to study pheromone signalling in *S. pombe*, have come out of the research group led by Graham Ladds (Smith *et al.*, 2009; Croft *et al.*, 2013).

The first model arose from the need to explain a highly repeatable experimental result that seemed to contradict the known biochemical function of Rgs1, namely, functioning as a GAP for the G protein Gpa1 (see subsection 1.3.8). It therefore came as a surprise that deletion of *rgs1*, a negative regulator, would decrease the maximum level of signalling output that yeast cells were capable of producing. However, in the absence of pheromone stimulation, the basal level of pathway activity was indeed elevated in *rgs1*Δ cells, in line with original expectations for this genotype.

Solving this conundrum required postulating a dual role for Rgs1, with both negative and positive effects on signalling output. The simplest way to assign a positive role for Rgs1 is to assume that GTP hydrolysis is somehow a limiting step for the activation of downstream effectors. This was translated into an extension

of the canonical G protein cycle (Yi *et al.*, 2003), where after a single round of downstream effector activation, Gα·GTP would transition into a hypothetical (still GTP-bound) inert state (Figure 3.1).

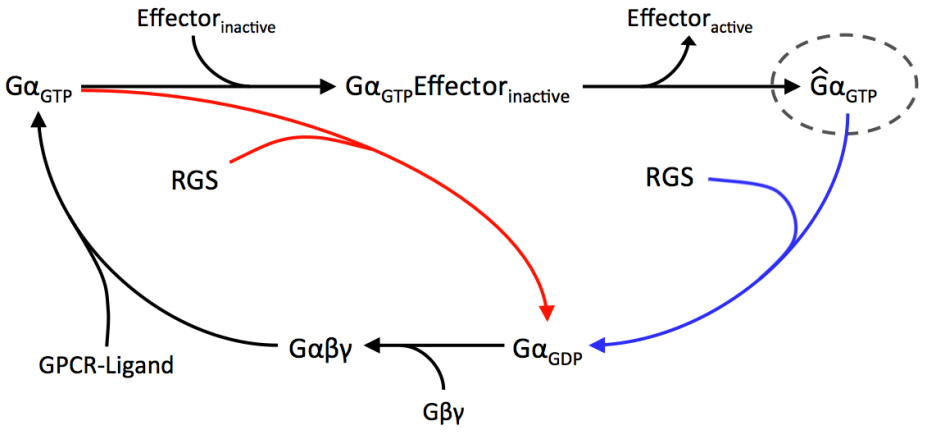


FIGURE 3.1: Inert Gα·GTP hypothesis. Accelerating GTP hydrolysis of Gα·GTP by Rgs1 has two different functions within the G protein cycle of *S. pombe* pheromone signalling. G protein deactivation before effector activation (red) limits the amount of unwanted pathway activation due to the intrinsic rate of Gpa1 nucleotide exchange. Ligand mediated GEF activity of the receptor overcomes the initial Rgs1 barrier and after activating one effector molecule, Gα·GTP enters an inert state (circled in dashed line), and GTP hydrolysis becomes rate-limiting for prolonged signalling. GAP activity of Rgs1 on inert Gα·GTP (blue) renews the availability of Gα·GDP to resume the cycle for as long as the receptor remains active.

This restructuring of the G protein cycle explained how deletion of *rgs1* would increase the basal levels of Gα·GTP due to spontaneous nucleotide exchange of Gpa1, and also provided a mechanism for Rgs1 being essential to reach full pathway activation, because once all Gpa1 molecules have become activated and thereafter inert, GTP hydrolysis is necessary to provide fresh input to restart the cycle. Importantly, the model proposed by Smith *et al.* (2009) predicted a counterintuitive relationship between Rgs1 concentration and maximal levels of signalling output that was confirmed experimentally (Figure 3.2) (Smith *et al.*, 2009).

Although highly praised by reviewers for its ability to predict unexpected levels of signalling output in response to experimental manipulation of Rgs1, the underlying assumption that allows the Smith *et al.* (2009) model to reproduce

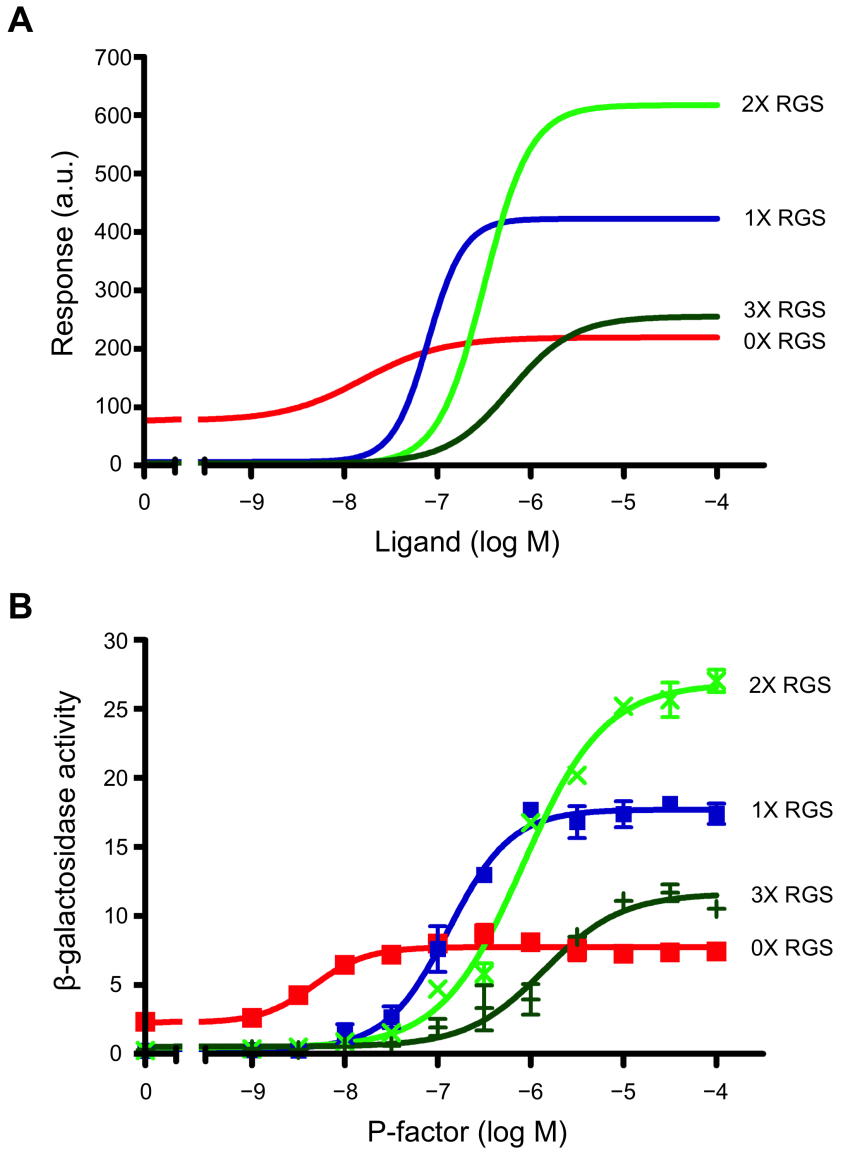


FIGURE 3.2: Dependence of pheromone signalling on Rgs1 concentration. Dose-response plots of total signalling output are compared between the mathematical model (A), and experiments (B) under four different Rgs1 concentrations relative to wildtype cells. Data were first published in Smith *et al.* (2009).

said experiments depends on a hypothetical inert Gα·GTP state. This assumption runs contrary to the accepted view of G protein function as a binary switch, and although the experiments could not be explained in any other way under a conventional G protein cycle, some reviewers were unwilling to entertain this hypothesis.

A subsequent project aimed to validate the existence of an inert Gα·GTP state

and to extend the model to include spatial regulation of Rgs1; however, Gpa1 turned out to lose all activity when tagged, despite a thorough attempt using many different epitopes inserted at various positions (Croft, 2012). This limitation hindered the ability to experimentally test the inert Gα·GTP hypothesis.

The work presented in this thesis began with the intention of using the Smith *et al.* (2009) model as a starting point upon which a whole systems view of pheromone signalling could be built, beyond the G protein cycle, that included feedback regulation at the gene network level. For this reason an immediate priority was attempting to recast the dual role of RGS into a more biologically plausible mechanism that would be easier to accept by other researchers in the field of G protein study.

3.1.1 Reformulating the inert Gα·GTP hypothesis

G proteins are signal transducers but they are not signal amplifiers, one Gα·GTP molecule will only activate one downstream effector until GTP hydrolysis terminates signalling (Purves *et al.*, 2004). The most well understood example of G protein signal transduction is that of Gα_s activating adenylate cyclase, where binding GTP creates a conformational change in Gα_s, allowing a protein-protein interaction that stabilizes the catalytically competent structure of adenylate cyclase (Tesmer *et al.*, 1997). In this case because adenylate cyclase is an enzyme, a long-lived interaction between Gα·GTP and effector will actually enhance the extent of signal transduction.

The direct effector that Gpa1 activates has not been identified, but it is known that between G protein and MAPK cascade, the signal must travel through the small GTPase Ras1 (Xu *et al.*, 1994). Crystallographic studies have revealed that Ras proteins use a single interface to interact with GEFs, GAPs, and downstream

effectors, which means there is competition between all possible interaction partners to bind Ras (Wittinghofer and Vetter, 2011). Thus, the lack of processivity by G proteins combined with competition for overlapping interaction surfaces in Ras, could create a potential bottleneck in signalling if Ras1 is in fact the direct effector activated by Gpa1.

In *S. pombe* there are two different GEFs that target Ras1, where activation by a particular GEF leads to downstream signalling of only one of two distinct pathways; however, details of how pathway specificity is achieved have not been established (Papadaki *et al.*, 2002; Tamanoi, 2011). Pheromone-dependent activation of Ras1 depends on the GEF Ste6, a cytosolic protein with no lipid modifications that would allow it to stably reside at the plasma membrane (PM), so like human RasGEFs, Ste6 probably requires a PM-bound adaptor protein to interact effectively with Ras1 (Hughes *et al.*, 1990; Papadaki *et al.*, 2002; Matsuyama *et al.*, 2006). An ideal candidate to fulfil this role is Gpa1, since it would immediately suggest how pathway-specific Ras activation is enforced. Furthermore, a proteome-wide interaction network of *S. pombe*, obtained by machine-learning using a variety of data that include conserved interactions across homologues, predicts with high confidence a direct interaction between Gpa1 and Ste6 (Pancaldi *et al.*, 2012).

Based on the information discussed above, a new reaction scheme for the G protein cycle of pheromone signalling can be proposed (Figure 3.3). Gpa1-dependent activation of Ras1 occurs in complex with Ste6. While bound to Gpa1, Ras1 is unable to interact with its effectors, so the half-life of the Gα·GTP-Effector complex can limit the rate of signal transduction. This complex would eventually break apart due to the intrinsic GTPase activity of Gpa1, but the rate of dissociation is greatly increased by GAP activity of Rgs1. In addition, activated Ras1 is subject to continual deactivation by its own GAP, Gap1 (Imai *et al.*, 1991).

This new G protein cycle is better grounded on established paradigms of G protein

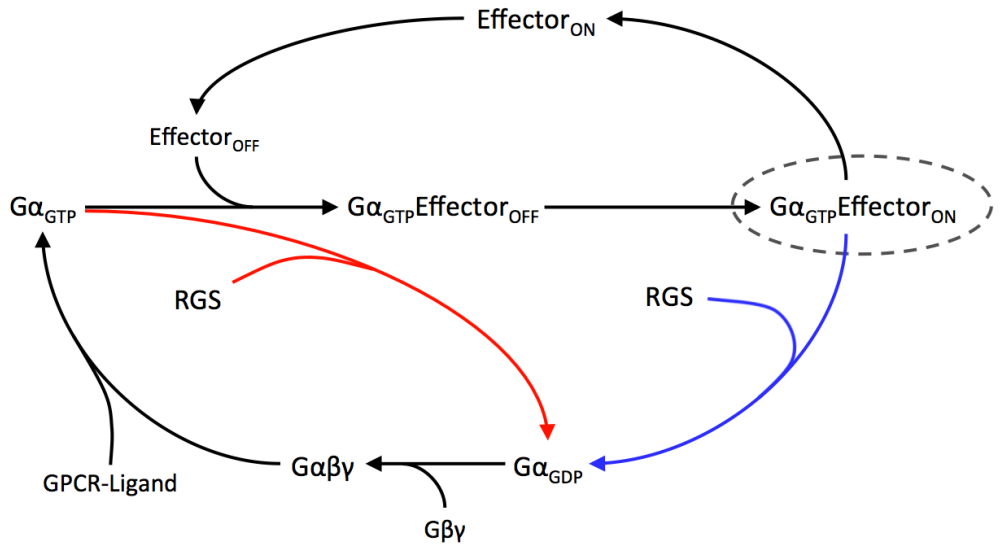


FIGURE 3.3: New G protein cycle proposed for *S. pombe* pheromone signalling. Rate-limiting complex dissociation is a more plausible mechanism than an inert Gα·GTP to explain the dual role of Rgs1 in pheromone signalling. RGS prevents Gα·GTP accumulation in the absence of receptor GEF activity (red). Gα·GTP can bind an inactive effector (OFF) and promote its transition towards an active state (ON). Further downstream signalling depends on release of the active effector which relies on GTP hydrolysis catalysed by RGS (blue). Without RGS the rate of complex dissociation will be reduced to the intrinsic GTPase activity of Gα, resulting in a reduced maximal rate of signalling output due to sequestration of effector (circled in dashed line).

biochemistry, and should be easier to validate experimentally compared to an inert Gα·GTP state. For example, by confirming the interaction between Gpa1 and Ste6 through co-immunoprecipitation, or by generating mutants that alter the lifetime of the complex and comparing the resulting dose-responses with model predictions. However, it can only be considered an acceptable replacement if it can successfully reproduce the Rgs1 perturbation experiments. A simulation analysis performed to compare the differences between the two reaction schemes found, as expected, that they are functionally equivalent and equally capable of explaining a dual role for Rgs1 in signalling (Figure 3.4).

As mentioned above, a separate project worked to extend the original Smith *et al.* (2009) model to include spatial regulation of Rgs1. Since that project was at a publishing stage when this project began, the model submitted for publication still had at its core the inert Gα·GTP assumption. In similar fashion to the previous

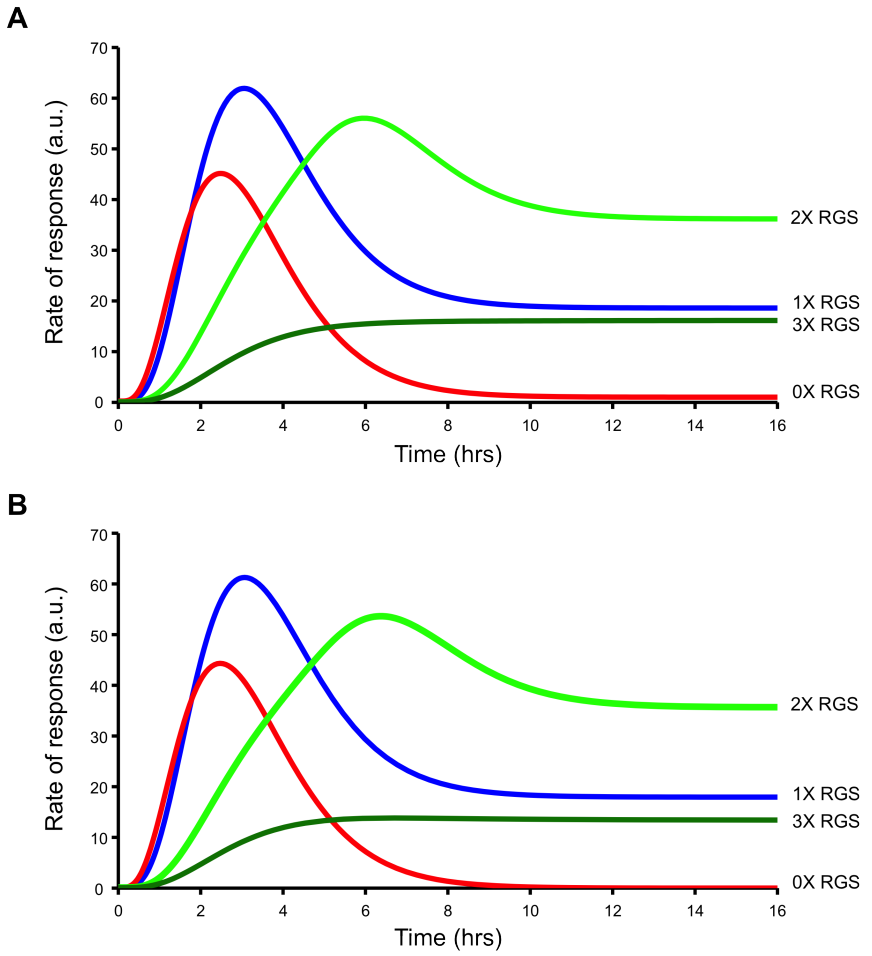


FIGURE 3.4: Equivalence between rate-limiting complex dissociation and an inert Gα·GTP state. The reaction schemes shown in figures 3.1 and 3.3 were simulated under identical conditions with four different RGS concentrations as in Figure 3.2. Plots show the rates of active effector production. (A) Simulation of the inert Gα·GTP model. (B) Simulation of the rate-limiting dissociation model.

case, publication was held back as reviewers challenged the basis of an inert Gα state. Shortly after having developed a better alternative, the extended model was rebuilt using the rate-limiting complex dissociation mechanism. This model was immediately accepted for publication (Croft *et al.*, 2013).

The initial work using the Smith *et al.* (2009) model highlighted several modelling considerations that made it inadequate for the goals of this project (section 1.5). The Smith *et al.* (2009) model was developed working with data that can only be compared qualitatively (transcriptional reporter), thus its parameters were not obtained through estimation from data, but rather defined arbitrarily. In addition,

many steps in the pheromone response pathway have not been described in detail, and so cannot be modelled explicitly to the same level of detail as in the Smith *et al.* (2009) model. To develop a quantitative model that allows investigating the roles of feedback control requires generating data that are appropriate for parameter estimation, and that the structure of the model ensures a high confidence in the estimated parameters, so that control can be correctly attributed to particular feedback loops. Under these constraints it is considered better practice to aim for a coarser model that includes less detail, but gains in confidence of the estimated parameters (Bornholdt, 2005; Raue *et al.*, 2011). For these reasons it was decided to develop a completely new minimal model of pheromone signalling that focuses on feedback regulation.

3.2 A minimal model of feedback control for pheromone signalling

The switch from proliferation to sexual differentiation in *S. pombe* depends on activation of the Ste11 transcription factor. All the mechanisms that regulate this decision converge their input on Ste11 to ensure its activation is timely and only occurs under appropriate circumstances (Anandhakumar *et al.*, 2013). Due to its central role in yeast mating, the focus of the model is placed on Ste11 and the feedback loops that are created around it.

Commitment to mating is mainly controlled by two signalling pathways, the nutrient-sensing cAMP pathway, and the pheromone response MAPK pathway (Otsubo and Yamamoto, 2012). For this reason it is common practice in studies that focus on pheromone signalling to use a *cyr1*Δ genotype as a base strain, to ensure that any observed effects are exclusively due to events controlled by the pheromone response (Xue-Franzén *et al.*, 2006). Yeast cells lacking the *cyr1* gene

have no detectable amounts of intracellular cAMP (Maeda *et al.*, 1990), which permanently derepresses Rst2-mediated *ste11* transcription (Higuchi *et al.*, 2002). Thus, *cyr1*Δ cells lose their dependence on nutrient-sensing for mating, and exhibit constitutive synthesis of both *ste11* mRNA and protein regardless of extracellular nutrient status (Sugimoto *et al.*, 1991; Valbuena and Moreno, 2010).

The work presented here adheres to this convention and utilised exclusively *cyr1*Δ yeast strains. This allows an initial simplifying assumption to be made and consider Ste11 protein as being continually produced at a fixed rate. In the absence of external perturbations the concentration of Ste11 will be in steady-state, balanced by degradation proportional to its own concentration (Figure 3.5). Another consequence of inactivating the cAMP pathway is that TORC1 becomes unable to block pheromone-induced nuclear localization of Ste11 under good nutritional conditions, even when Tor2 is overexpressed (Valbuena and Moreno, 2010). For this reason any effects of the TORC1 pathway are ignored in this model.

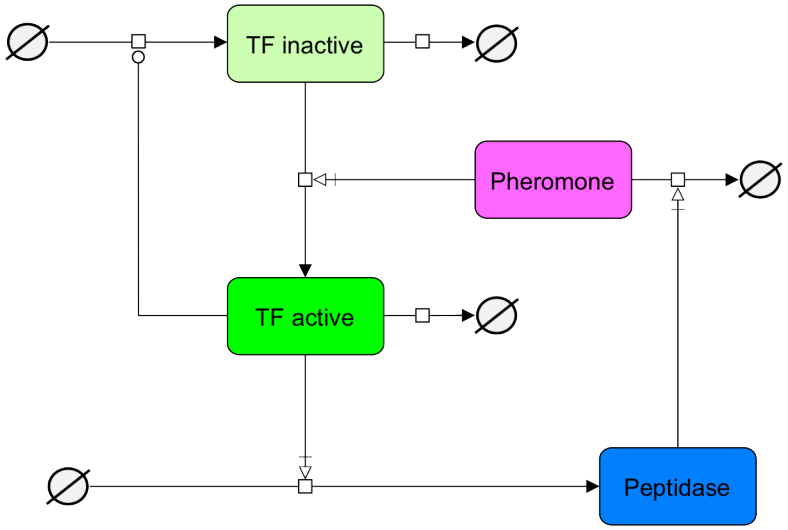


FIGURE 3.5: Model diagram constructed using Systems Biology Graphical Notation (SBGN) (Kitano *et al.*, 2005). State variables are represented by labelled rectangles. Transformation reactions are shown with black head arrows. White head arrows affecting reactions indicate that the state variable is a necessary modifier for the reaction to occur. White circled connectors indicate positive modulation. Sources and sinks are represented by null set symbols.

Newly synthesized Ste11 is prevented from activating its target genes through nuclear exclusion by the highly abundant cytosolic Rad24 protein (Kitamura *et al.*, 2001; Marguerat *et al.*, 2012). Pheromone stimulation relieves Rad24 inhibition by MAPK phosphorylation of Ste11, allowing Ste11 to dissociate from Rad24 and accumulate in the nucleus (Qin *et al.*, 2003). The model accounts for this regulation by considering all production of Ste11 into an inactive pool. The inactive transcription factor becomes activated by injection of pheromone, where activation occurs at a rate proportional to the concentration of both pheromone dose and transcription factor remaining in the inactive pool.

Activated Ste11 is the master regulator that triggers the initiation of the mating genetic programme, which includes dozens of genes required for specific stages of the mating process such as cell fusion, nuclear movement, and meiosis (Mata and Bähler, 2006; Mata *et al.*, 2007). Among the earliest expressed genes in this transcriptional network is the serine carboxypeptidase Sxa2, as well as upregulation of the *ste11* gene itself (Xue-Franzén *et al.*, 2006). Sxa2 is exported into the extracellular space where it inactivates P-factor pheromone (Ladds *et al.*, 1996; Ladds and Davey, 2000), thus, the transactivation of *ste11* and *sxa2* in response to pheromone stimulation gives rise to a pair of coupled positive and negative feedback loops that are built-in directly into the structure of the network. The strength of the feedback loops and their impact on signalling dynamics will be determined by the reaction rates of each step in this circuit. All of these rates can be considered in the model through parameters that can be estimated by fitting the model to data. The corresponding dynamic equations for the complete model

structure presented in Figure 3.5 are given by

$$\frac{d}{dt}\,x_1 = k_b \,+\, v_1\,\frac{x_2}{k_{m_1} + x_2} \,-\, k_d \cdot x_1 \,-\, k_a \cdot x_1 \cdot x_4 \tag{3.1}$$

$$\frac{d}{dt}\,x_2 = k_a \cdot x_1 \cdot x_4 \,-\, k_d \cdot x_2 \tag{3.2}$$

$$\frac{d}{dt}\,x_3 = v_2\,\frac{x_2}{k_{m_2} + x_2} \tag{3.3}$$

$$\frac{d}{dt}\,x_4 = -\,k_c \cdot x_3 \cdot \frac{x_4}{k_{m_3} + x_4} \tag{3.4}$$

with initial condition vector

$$\boldsymbol{x}(0) = \boldsymbol{x_0}(\boldsymbol{p}) = \begin{bmatrix} \dfrac{k_b}{k_d} & 0 & 0 & d \end{bmatrix}^{\mathsf{T}} \tag{3.5}$$

where $\boldsymbol{x}(t) = [x_1, x_2, x_3, x_4]^{\mathsf{T}}$. The definitions of the state variables and parameters for the model (3.1)-(3.5) are given in Table 3.1.

3.2.1 Identifying areas of uncertainty in the model assumptions

Despite limiting the initial model derivation to a minimum of state variables, there are still a number of alternative descriptions that could be chosen to represent this system. These alternatives arise because of the need to choose a particular level of detail for each process, where multi-step events can be modelled explicitly by their elementary reactions, or implicitly by lumping steps into a single reaction that preserves the same input-output velocity (Klipp *et al.*, 2009; Ay and Arnosti, 2011). Another source of alternatives comes from basic lack of knowledge about certain steps in the pathway, which introduces uncertainty as to whether the chosen kinetics are appropriate for modelling a given reaction (Alves *et al.*, 2008). Finally,

TABLE 3.1: Definitions of model A1B1C1D1E1 presented in equations (3.1)-(3.5). Concentration units are in nM. Time units are in min.

State Variables	Description	Dimensions†
x_1	Transcription factor inactive (Ste11)	$\mathcal{C}$
x_2	Transcription factor active (Ste11)	$\mathcal{C}$
x_3	Peptidase (Sxa2)	$\mathcal{C}$
x_4	Pheromone (P-factor)	$\mathcal{C}$
Parameters		
k_b	Basal rate of TFi	$\mathcal{C}\,\mathcal{T}^{-1}$
v_1	Max. rate of TFi production	$\mathcal{C}\,\mathcal{T}^{-1}$
k_{m_1}	Michaelis-Menten constant	$\mathcal{C}$
k_d	TF degradation constant	$\mathcal{T}^{-1}$
k_a	TF activation constant	$\mathcal{C}^{-1}\,\mathcal{T}^{-1}$
v_2	Max. rate of Peptidase production	$\mathcal{C}\,\mathcal{T}^{-1}$
k_{m_2}	Michaelis-Menten constant	$\mathcal{C}$
k_c	Peptidase catalytic constant	$\mathcal{T}^{-1}$
k_{m_3}	Michaelis-Menten constant	$\mathcal{C}$
d	Experimentally controlled dose	$\mathcal{C}$

† where $\mathcal{C}$ and $\mathcal{T}$ denote concentration and time respectively.

it may be desirable to build alternative models simply to evaluate the importance of including specific assumptions to adequately explain the data.

With these considerations in mind, the original model assumptions were evaluated according to the confidence that can be ascribed to each one, and those responsible for the greatest uncertainty were selected to generate a family of related models, all sharing the same feedback structure as in Figure 3.5. To simplify the referencing of specific model variants an identifier nomenclature is introduced where a letter code represents a model assumption and a number after each letter represents

the option chosen to model each variant. The model variants along with their corresponding codes are summarised in Table 3.2. According to this code the base model introduced in equations (3.1)-(3.5) would have the identifier A1B1C1D1E1.

Table 3.2: Model variants.

Identifier	Description	Options	Implementation
A	Transcription and translation described in one or two steps	1 2	One-step model Two-step model
B	Activation of TF is fast or occurs after a delay	1 2	No delay Delay
C	Degradation of active / inactive TF has the same or different rate	1 2	Same rate Different rate
D	Transcription Hill coefficient is fixed or a free parameter	1 2	Fixed Hill coefficient Free Hill coefficient
E	Positive feedback	1 2	Included Not included

3.2.2 Model granularity

During the initial stage of the modelling process the best option for model granularity may not be immediately obvious if the types of data that will be used for parameter estimation are not exactly known. There is always a trade-off between increasing the level of detail in a model and the cost it generates in the number of free parameters. Models of gene regulatory networks typically choose between two main alternatives for their level of detail describing gene expression: 1) models may consider transcription and translation as a single reaction as in equations (3.1) and (3.3); or 2) the dynamics of mRNA and protein can be considered separately

in a two-step process (Klipp *et al.*, 2009), namely:

$$\frac{d}{dt}\,mRNA = v_{max}\,\frac{TF}{K_m + TF} \;-\; k_{dm} \cdot mRNA \tag{3.6}$$

$$\frac{d}{dt}\,protein = k_t \cdot mRNA \;-\; k_{dp} \cdot protein \tag{3.7}$$

In this model, transcription is typically described with Michaelis-Menten or Hill-type kinetics, and translation follows mass-action kinetics according to the concentration of mRNA. In addition, both mRNA and protein are degraded with mass-action kinetics. Overall, the cost of including one more species in the two-step model results in two extra parameters to be fitted for each gene compared to the one-step model of gene expression.

3.2.3 Appropriateness of reaction kinetics

Genetic and biochemical experiments have firmly established the chain of causality for the structure shown in Figure 3.5 (see section 1.3.8), however, several details with respect to specific timing of signalling events have remained unexplored. Most notably, the time taken between pheromone detection by membrane-bound receptors (Mam2/Map3) and Ste11 activation has not been determined. The assumption behind equation (3.2) is that signalling is fast compared to the time scale of gene expression, and thus activation of inactive Ste11 molecules begins immediately after pheromone stimulation. Alternatively, if signalling between receptor and transcription factor takes an appreciable amount of time, a generic time delay

of the following form may be inserted between these events:

$$\frac{d}{dt}\,delay_1 = k_{delay} \cdot TF_{inactive} \; - \; k_{delay} \cdot delay_1 \tag{3.8}$$

$$\vdots$$

$$\frac{d}{dt}\,delay_n = k_{delay} \cdot delay_{n-1} \; - \; k_{delay} \cdot delay_n \tag{3.9}$$

$$\frac{d}{dt}\,TF_{active} = k_{delay} \cdot delay_n \; - \; k_{deg} \cdot TF_{active} \tag{3.10}$$

This time delay would then capture the correct activation dynamics Ste11 without needing to explicitly describe all the intermediary steps for which there are no experimental readouts that can justify their inclusion. Although the time delay suggested in equations (3.8)-(3.10) is not a true time delay in the sense of time-delay systems (Driver, 1977), this form is much simpler for simulation and analysis, and as such is much more commonly used in systems biology models (Maiwald and Timmer, 2008).

Another process with uncertain timing in the model is the degradation of transcription factor. Upper bounds have been determined for Ste11 degradation in the absence of pheromone (Kjærulff *et al.*, 2007), however, it is not known whether MAPK phosphorylation or shuttling from cytosol to nucleus have any effect on the speed of Ste11 degradation. In the event that degradation rates are distinct between these two species, equations (3.1)-(3.2) can be modified to fit an independent parameter for each one, with the original degradation parameter k_d being split into k_{d1} for TF-inactive, and k_{d2} for TF-active.

Finally, the biophysics of Ste11-mediated transcription remain unknown. The precise mechanism by which Ste11 interacts with target promoters and the general transcription machinery can have relevant consequences for transcriptional kinetics, particularly if several Ste11 molecules can cooperate to facilitate transcription

initiation. Theoretical derivations of deterministic transcription rates result in saturation kinetics (Klipp *et al.*, 2009), such as Michaelis-Menten kinetics as used in equations (3.1) and (3.3), however, this description represents a specific instance of the more general case of the Hill equation, which includes a parameter, the Hill coefficient, that accounts for cooperative effects:

$$v_{max} \frac{x^n}{K_m^n + x^n} \tag{3.11}$$

where n is the Hill coefficient. The Michaelis-Menten case can be thought of as having fixed the Hill coefficient to a value of 1. Hill kinetics are important when describing switch-like responses, for example if the transition of a promoter from a silent state into maximum transcription occurs at a very steep rate. Although the Hill equation is a more comprehensive description of transcription, estimating Hill coefficients accurately usually requires highly resolved time series data, so often the choice is made to use the simpler Michaelis-Menten option (Raue *et al.*, 2013).

3.2.4 Importance of positive feedback

An implicit assumption of the model (3.1)-(3.5) is that positive feedback from an activated transcription factor to increase the rate of its own expression is an important feature of the system that must be included to understand its behaviour. Although the existence of this feedback is suggested based on experimental evidence (Kunitomo *et al.*, 2000; Xue-Franzén *et al.*, 2006), unlike the negative feedback loop of the model, the role of positive feedback in pheromone signalling has not been investigated (Ladds *et al.*, 1996; Ladds and Davey, 2000). A first step towards understanding the role of this positive feedback can be made by comparing the dynamics of models with and without this feedback loop, and assessing its importance towards explaining experimental data. A model variant without

positive feedback can be made by replacing equation (3.1) with

$$\frac{d}{dt}\, x_1 = k_b \; - \; k_d \cdot x_1 \; - \; k_a \cdot x_1 \cdot x_4 \tag{3.12}$$

where the production of TF-inactive depends exclusively on the basal rate of transcription by Rst2, k_b.

3.3 Initial model assessment

The model derivation from existing knowledge in section 3.2 resulted in a family of 32 model variants related to each other by the network structure shown in Figure 3.5. Distinguishing the adequacy of these models to describe the dynamics of pheromone signalling requires fitting the models to experimental data to assess their relative ability to capture the system behaviour. Based on the state variables considered in the models, there is only one published data set that can be used for model fitting. Ladds *et al.* (1996) reported the time-dependent velocity of P-factor inactivation by Sxa2 obtained from samples of growth medium where yeast cells had been exposed to various concentrations of pheromone (Figure 3.6).

Assuming that the reaction velocity is linearly proportional to Sxa2 concentration, as in equation (3.4), the data generated by Ladds *et al.* (1996) are related to the models by the following output structure:

$$y^{\dagger} = S_{pep} * Peptidase \tag{3.13}$$

where $y^{\dagger}$ are the experimental measurements, and 'Peptidase' the corresponding state variable in the models. A scaling parameter S_{pep} is introduced to account for the conversion between the units of percentage activity and concentration.

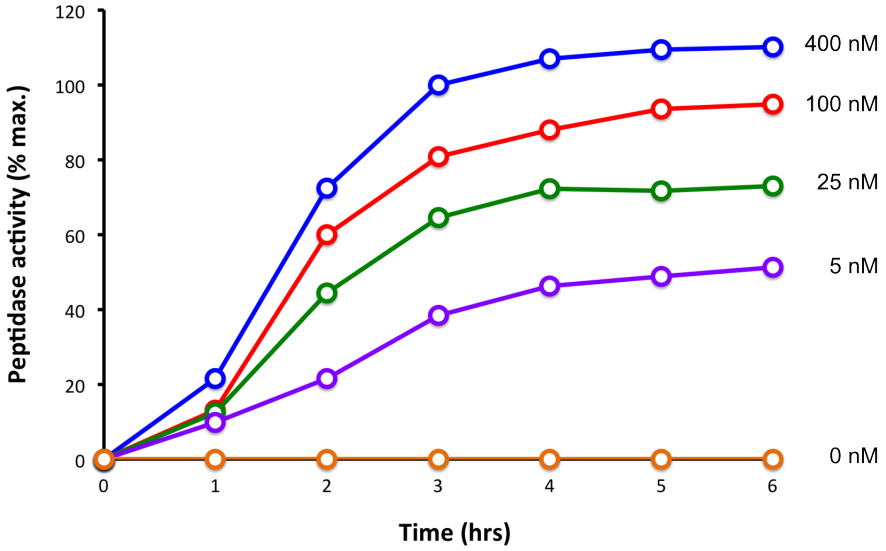


FIGURE 3.6: Sxa2 production in response to P-factor stimulation. Mid-exponential cultures ($\sim 5 \times 10^6$ cells $mL^{-}1$) of M *cyr1*$^{-}$ cells were treated with different concentrations of P-factor, growth medium samples were collected every hour and assayed immediately. The medium was filtered to remove cells and supplemented with P-factor to a final concentration of 12 µM. Following P-factor addition 1 mL aliquots were withdrawn from the test sample at small intervals, adjusted to 0.5% TFA to stop the reaction, and the conversion of active to inactive P-factor (ΔLeu) was analysed by reversed-phase HPLC on a C_{18} column using an acetonitrile gradient. Sxa2 activity is expressed as a percentage of the reaction rate observed with medium prepared 3 h after treating cells with P-factor at 400 nM.

Estimation of the model parameters was done by minimizing the weighted sum of squared residuals

$$\chi^2(\theta) = \sum_{k=1}^{m} \sum_{i=1}^{d_k} \left(\frac{y_{ki}^{\dagger} - y_k(t_i, \theta)}{\sigma_{ki}} \right)^2 \tag{3.14}$$

where the experimental data $y_{ki}^{\dagger}$ for each observable, $k = 1$ to m, are measured at time points t_i with $i = 1$ to d_k, with measurement error σ_{ki}. All parameters are by definition positive definite, $\theta \in \mathbb{R}^+$, so estimation is performed in log-space, which also allows efficient numerical handling of large differences in parameters that are potentially orders of magnitude apart. The optimization problem was solved with a deterministic nonlinear least squares algorithm with user-supplied derivatives as recommended by Raue *et al.* (2013).

To give each model the best opportunity of fitting the data, a multi-start approach

with Latin hypercube sampling (LHS) was employed to choose initial parameter values. LHS ensures that the multidimensional parameter space is sampled more efficiently than random sampling, and that each independent optimization run starts in a different region of the search space (Owen, 1992). For practical purposes this strategy converts the selected gradient search algorithm into a global optimizer (Raue and Timmer, 2011). The number of fits required can be chosen by looking at the value of the objective function at the end of each run and sorting all runs according to their likelihood (Figure 3.7). A high performance method should exhibit a step-wise arrangement, indicating that a local optimum was reached in every case. In contrast, low performance methods will display a continuous slope between best and worst fits, which suggests that the search was stopped at a suboptimal solution (Raue *et al.*, 2013). For the size and complexity of the models analysed here 100 runs were more than enough to reliably reach the best possible fit in each case.

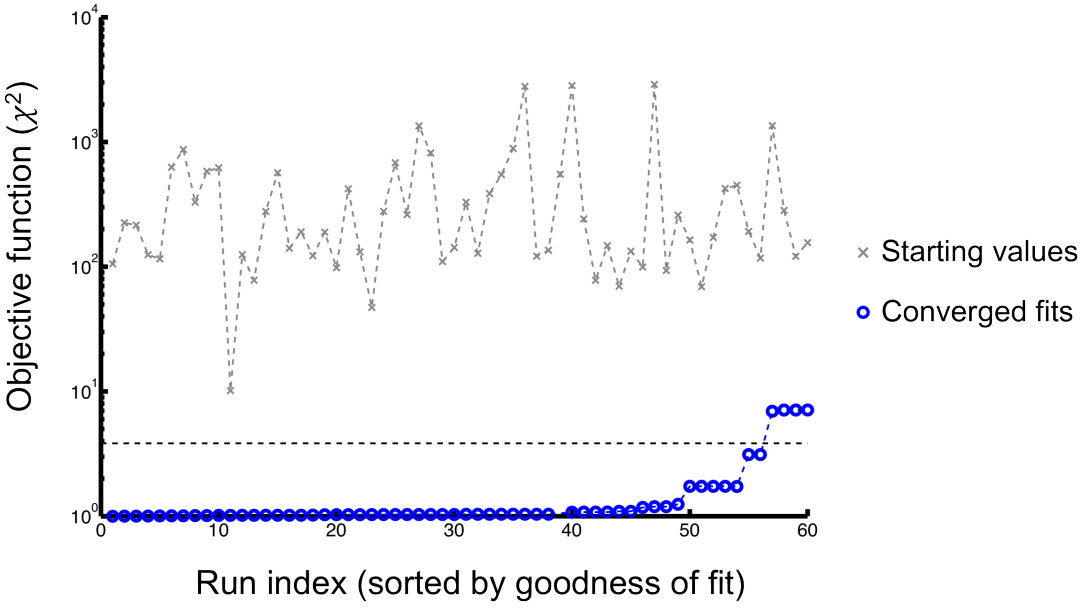


FIGURE 3.7: Performance of multi-start parameter estimation. A representative example of fitting one of the 32 model variants to the Ladds *et al.* (1996) data is shown, with 60 independent optimization runs that achieved successful convergence sorted by decreasing goodness of fit (blue circles). The starting χ^2 value of the objective function is shown for each run above its corresponding converged fit (grey x marks). χ^2 values that fall below the dotted line represent solutions with a qualitatively good fit.

After successful parameter estimation was achieved for all model variants, most models were able to give a sensible fit to the data. In order to discriminate between the models in an unbiased manner there are two main strategies commonly employed for model selection. The first alternative is to use a statistical test, where the null hypothesis states that two competing models are equally good at fitting the data and the simpler model is favoured by default. The null hypothesis is rejected with confidence level alpha if the more complex model fits the data significantly better. The second alternative for model selection is to use a selection criterion, where a scoring function is used to calculate a measure that balances between a better agreement with experimental data and the number of extra parameters required for the better fit (Motulsky and Christopoulos, 2004; Klipp *et al.*, 2009).

Statistical hypothesis testing has the limitation that it can only be used between nested models, i.e. one of the models must be a subset of the other. In our case there is a mixture of nested and non-nested models, for example, D1 models are a subset of D2 models (fixed vs free Hill coefficients), but A1 models are not a subset of A2 models (one-step vs two-step gene expression). In addition each hypothesis test must be done on only two models at a time, which would require 992 pairwise tests to compare all models against each other, giving a high probability of reaching false conclusions in many cases. In contrast, selection criteria can be used to compare non-nested models, and can be used for simultaneous comparison of all models at once. Selection criteria do not use an arbitrary threshold to reject models; instead of just stating which model is more likely, they also give information about how much more likely one model is over another (Motulsky and Christopoulos, 2004).

Here we use the Akaike Information Criterion (AIC), originally derived from principles of information theory (Akaike, 1974). The AIC rewards goodness-of-fit as described by the likelihood function, while directly penalising the number of free parameters that are estimated. In general models with lower AIC values should

be preferred over models with higher values; however, the AIC does not give direct information about the quality of a fit, so these must be assessed independently in the context of the problem. The AIC score for a model fitted to a particular dataset is given by

$$\mathrm{AIC} = 2\,\mathrm{p} - 2\ln\left(L\left(\theta\right)\right) \tag{3.15}$$

where p is the number of parameters estimated, and where

$$L(\theta) = \prod_{k=1}^{m}\prod_{i=1}^{d_k}\frac{1}{\sqrt{2\pi\sigma_{ki}^2}}\exp\left(-\frac{1}{2}\left(\frac{y_{ki}^{\dagger} - y_k(t_i,\theta)}{\sigma_{ki}}\right)^2\right) \tag{3.16}$$

is the likelihood function for a given parameter set θ. To compute the AIC we note that for normally distributed measurement error, the least squares problem of obtaining parameter estimates ($\hat{\theta}$), i.e.

$$\hat{\theta} = \arg\min\left[\chi^2(\theta)\right] \tag{3.17}$$

is equal to maximum likelihood estimation (MLE) through the relationship

$$\chi^2(\theta) = \mathrm{C} - 2\ln(L(\theta)) \tag{3.18}$$

for a suitable constant C. In the case where the number of data points is small with respect to the number of parameters that need to be estimated, the standard AIC will not be stringent enough against the addition of extra parameters and requires a bias correction. The corrected AIC (AICc) is given by

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2\,\mathrm{p}\,(\mathrm{p}+1)}{\mathrm{N}-\mathrm{p}-1} \tag{3.19}$$

for N data points and p estimated parameters. Since both the AIC and AICc converge for large N, in general AICc should always be used in practice (Burnham and Anderson, 2002). For readability, from here-on we refer to the AICc simply

as the AIC.

After ranking all the models according to their AIC score we can select the best models for further analysis by focusing on one model assumption at a time (Table 3.2). In each case, for a particular model assumption there will be 16 different models that follow option 1, and 16 matching models that are identical to the first 16 except in following option 2. This allows evaluating the effects that each assumption has on the AIC score several times with different pairs of directly comparable models.

A2 models (two-step gene expression) display a consistently higher AIC than their A1 counterparts when fitted to the Ladds *et al.* (1996) data (Figure 3.8 A). This is a clear indication that this assumption generates models that are too complex for the amount of information in this data set, because the two extra parameters required by A2 models do not translate into a significantly better fit compared to A1 models (Figure 3.8 B). To avoid confounding the analysis of further assumptions, A2 models were excluded from the rest of this initial assessment.

From Figure 3.8 B it is evident that the choice of option E can have a large impact on model fitting, as a pair of E2 models (no positive feedback) standout as outliers when comparing χ^2 values; for this reason this assumption was analysed next. Looking at the differences in AIC between E1 and E2 models, it appears at first that E2 models would be preferable because in general they produce lower AIC scores (Figure 3.9 A); however, cross-checking with the χ^2 values shows that E2 models have trouble achieving a good fit to the data (Figure 3.9 B). Counting the number of parameters in each model reveals that having at least 9 parameters is a minimum requirement to achieve a satisfactory visual fit with these models (Figure 3.9 C). This suggests that the observed trend of lower AIC values in E2 models is an artifact caused solely by the reduced number of parameters in these

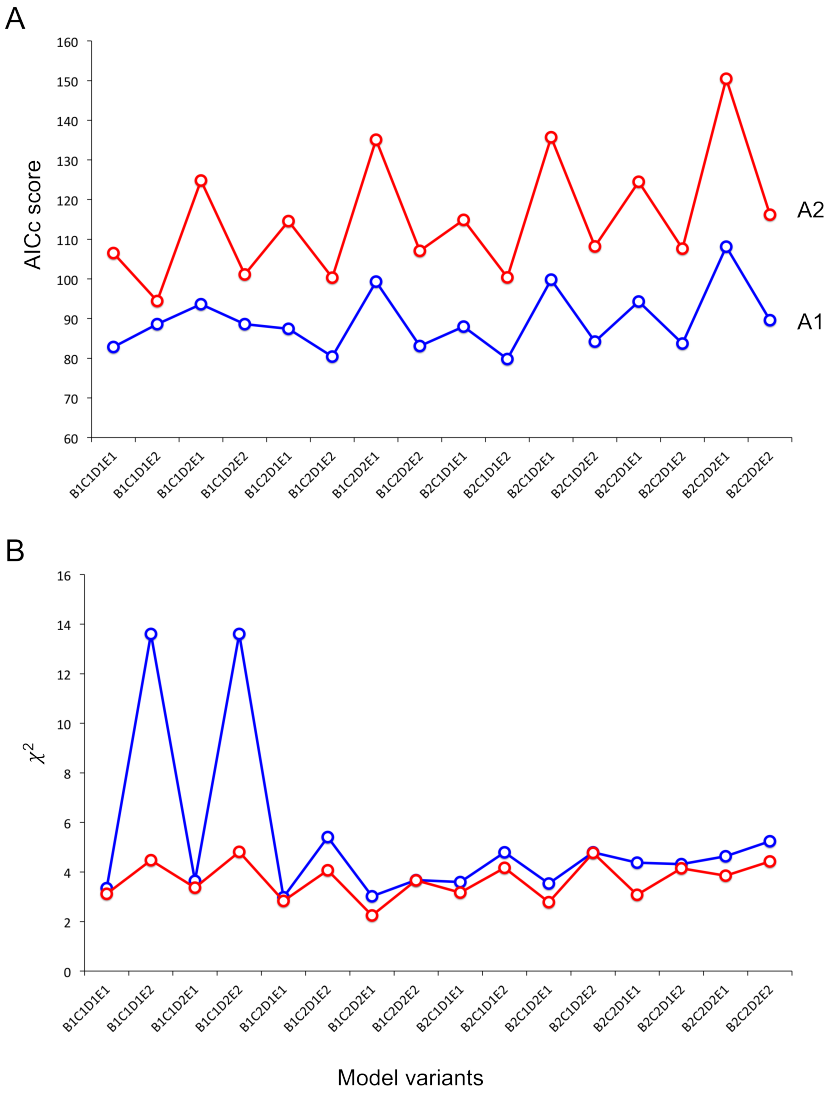


FIGURE 3.8: Comparison of fits between models of class A1 (blue) and A2 (red). (A) AIC scores. (B) χ^2 values.

models, and does not indicate that lack of positive feedback is a better assumption to explain the data.

Based on goodness-of-fit, all but one of the E2 models can be discarded. Model A1B1C2D2E2 is directly comparable to the best AIC ranking E1 model, which is the original base model A1B1C1D1E1. The E2 variant compensates the lack of positive feedback on transcription factor production by the addition of parameters in other areas that can affect its concentration, e.g. degradation rates in assumption C; however, the E1 variant is still able to reach a moderately lower χ^2 value (3.34 vs 3.67), which translates to a better AIC score by just 1 point

(Figure 3.9 A). The significance of this difference can be assessed by the evidence ratio (Motulsky and Christopoulos, 2004), which is simply the relative likelihood of having chosen the correct model given by

$$\text{Evidence ratio} = \frac{\text{P model 1 is correct}}{\text{P model 2 is correct}} = \frac{1}{e^{-0.5\Delta\text{AIC}}} \tag{3.20}$$

where ΔAIC is the difference in AIC scores between the models. Computing the evidence ratio for the E model variants shows that the E1 variant is $\sim$1.2 times more likely to be correct, or 55% chance, which is a negligible difference. This information alone is not sufficient to make any judgements about the importance of including or excluding a positive feedback loop in the model structure; however, because option E2 does in general produce unsatisfactory fits and there are published reports supporting the existence of this feedback (Kunitomo *et al.*, 2000; Xue-Franzén *et al.*, 2006), option E1 was initially considered as the default option.

Analysis of model assumptions B and C showed similar results, where in every case the option with less parameters exhibited a lower AIC, with no difference observed in χ^2 goodness-of-fit between the simpler and more complex models. After controlling for the effect of minimum number of parameters, this represents the same situation encountered for assumption A (Figure 3.8), where the data set used for parameter estimation lacks enough information to discriminate between model variants.

Assumption D appeared at first to also fall into the category of assumptions that could not be identified, where having the Hill coefficient as a free parameter did not improve goodness-of-fit; however, looking at the combined results of multiple parameter estimation runs it became apparent that the optimization procedure consistently converged on a Hill coefficient of 1 for both gene expression terms in equations (3.1) and (3.3), which suggested there could be distinguishable effects between the two options that would allow a selection decision to be made.

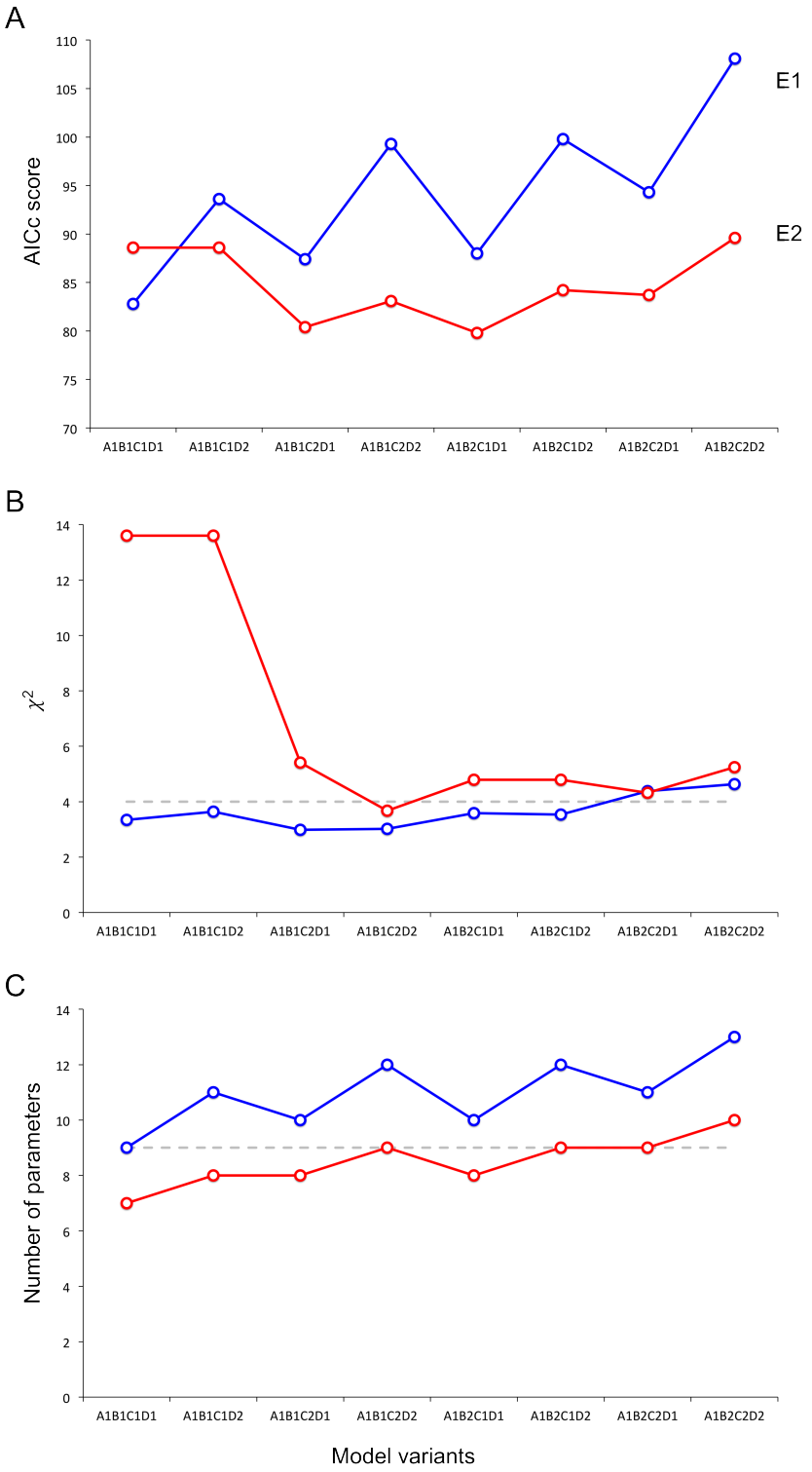


FIGURE 3.9: Comparison of fits between models of class E1 and E2. (A) AIC scores. (B) χ^2 values. (C) Number of parameters in each model. Grey dotted line shows the threshold for a qualitatively good fit.

To visualize the effects of varying the Hill coefficients, a parameter scan was performed for both gene expression terms. Changing the Hill coefficient n_1 affecting

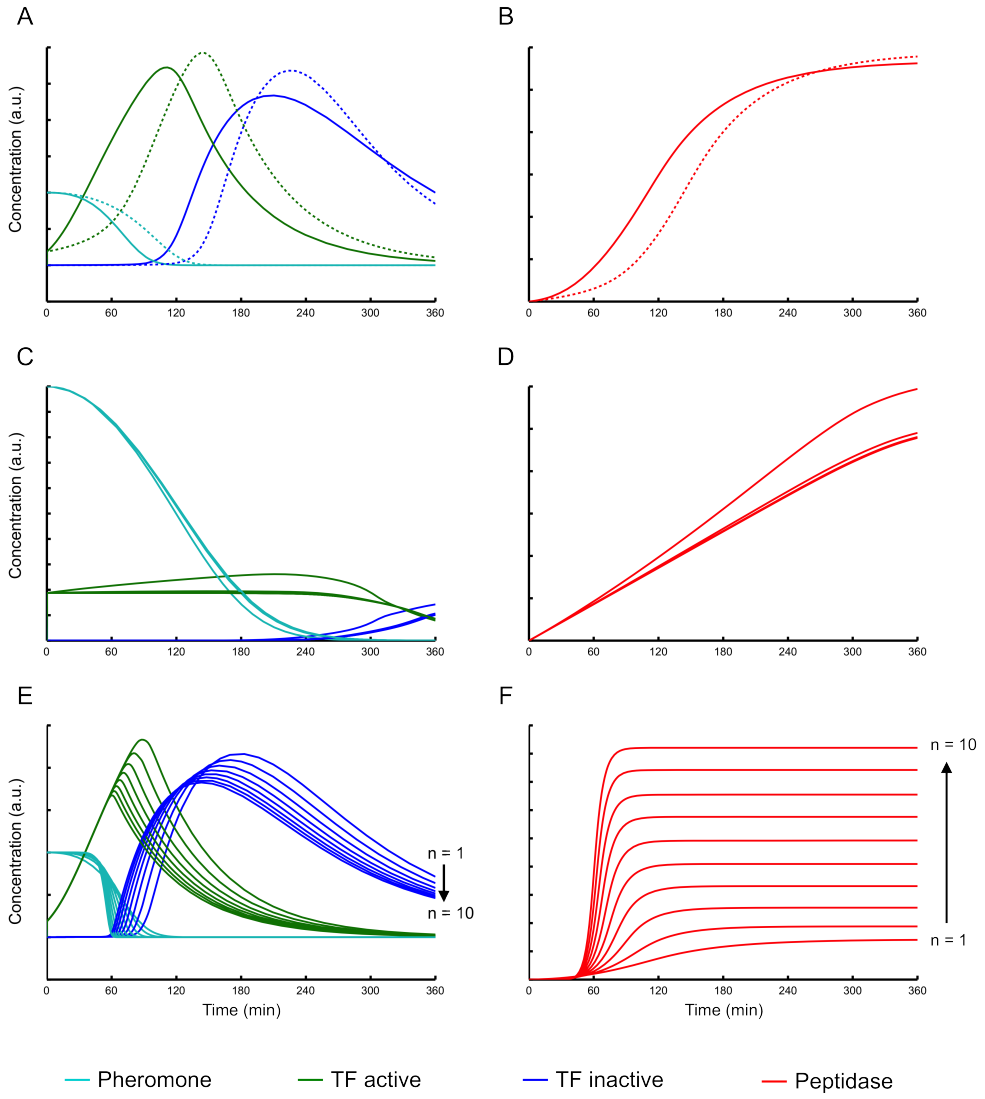


FIGURE 3.10: Identification of Hill coefficients for gene expression terms. A parameter scan was performed for Hill coefficients involved in TF production, n_1, and Peptidase production, n_2. (A, B) Values of $n_1 = 1$ (solid lines) and $n_1 = 2$ (dashed lines) both preserve the sigmoidal shape required to fit the Ladds *et al.* (1996) data, but they exhibit differences along the time axis that allows discrimination between them. (C, D) Values of $n_1 > 2$ can be readily dismissed because they alter the shape of Peptidase production into a straight line. (E, F) Varying the value of n_2 changes the steady-state concentration of Peptidase reached after pheromone stimulation.

transcription factor production in equation (3.1) produced a sharp transition in model behaviour at the boundary between $n_1 = 2$ and $n_1 = 3$ (Figure 3.10). For $n_1 > 2$ the state variables become insensitive to the value of n_1 and converge on the same simulation trajectories (Figure 3.10 C, D); more importantly however,

the shape of Peptidase production becomes linear, and this effect cannot be compensated by other model parameters. This allows values of $n_1 > 2$ to be discarded as incompatible with the Sxa2 production data. For $n_1 \leq 2$ the simulation of Peptidase production retains a sigmoidal shape, but the differences are large enough that a value of $n_1 = 1$ can be selected with confidence during the estimation procedure (Figure 3.10 A, B). Further confirmation of this value could be obtained from measurements of additional state variables which also show timing differences in their trajectories between $n_1 = 1$ and $n_1 = 2$.

Manipulating the Hill coefficient n_2 affecting Peptidase expression in equation (3.3) translated into smooth changes in model behaviour rather than any sharp transitions (Figure 3.10 E, F). The main effect of increasing n_2 on Peptidase production was that it raised the maximum level of Peptidase concentration reached at steady-state. Since the data from Ladds *et al.* (1996) are not on an absolute scale this number cannot be fixed; however, values of Hill coefficients other than $n_2 = 1$ reduced the necessary flexibility in the model to simultaneously fit the four different doses in the data set, leading again to a convergent estimation of $n_2 = 1$. Overall, this allows the option D1 (fixed Hill coefficients) to be selected for further development of the model.

3.4 Parameter determinability

Screening the model candidates through parameter estimation in section 3.3 allowed several unsuitable variants to be discarded; however, the analysis also highlighted the importance of matching the complexity of the models to the information density contained in the data sets used for model calibration. Although the data from Ladds *et al.* (1996) contains time-resolved response measurements for multiple doses, many of the modelling assumptions could not be discriminated by this data set alone. The best combination of AIC score and χ^2 goodness-of-fit

value was obtained by the base model A1B1C1D1E1 (Figure 3.9), which appears to have the correct size for fitting to the Ladds *et al.* (1996) data. This model has the least number of parameters from among the models that were able to achieve satisfactory fits. After fitted to experimental data one of the most important uses for a model is to explore the dynamics of unobserved states through simulation. By simulating model A1B1C1D1E1 using the parameter estimates obtained from its best fit, several predictions can be generated about the behaviour of the pheromone signalling pathway (Figure 3.11). P-factor pheromone is expected to be completely inactivated several hours before the production of Sxa2 peptidase finally stops (Figure 3.11 B). Also, the concentration of TF-active appears to increase rapidly after pheromone stimulation, reaching a peak level that does not increase linearly with pheromone dose (Figure 3.11 C), suggesting there could be a threshold in levels of TF-active that dictate commitment to sexual differentiation. However, before any meaning can be extracted from the simulations, it must be assessed whether the parameters have been determined from the estimation procedure with reasonable confidence. Since the model trajectories depend on the estimated parameters, any uncertainty in the parameters translates directly into uncertainty in the simulations.

Practical identifiability analysis (see section 1.4.1) can be performed through several methods (Hengl *et al.*, 2007; Balsa-Canto and Banga, 2010; Raue *et al.*, 2011). The simplest way of discovering hidden dependencies between the parameters is by checking their cross-correlation κ (Jaqaman and Danuser, 2006), where for two parameters i and j

$$\kappa_{ij} = \frac{\hat{V}_{ij}}{\sqrt{\hat{V}_{ii}}\sqrt{\hat{V}_{jj}}} \tag{3.21}$$

with $\hat{V}$ being the variance-covariance matrix of the estimated parameters. A symmetric correlation matrix can be constructed with the normalized absolute

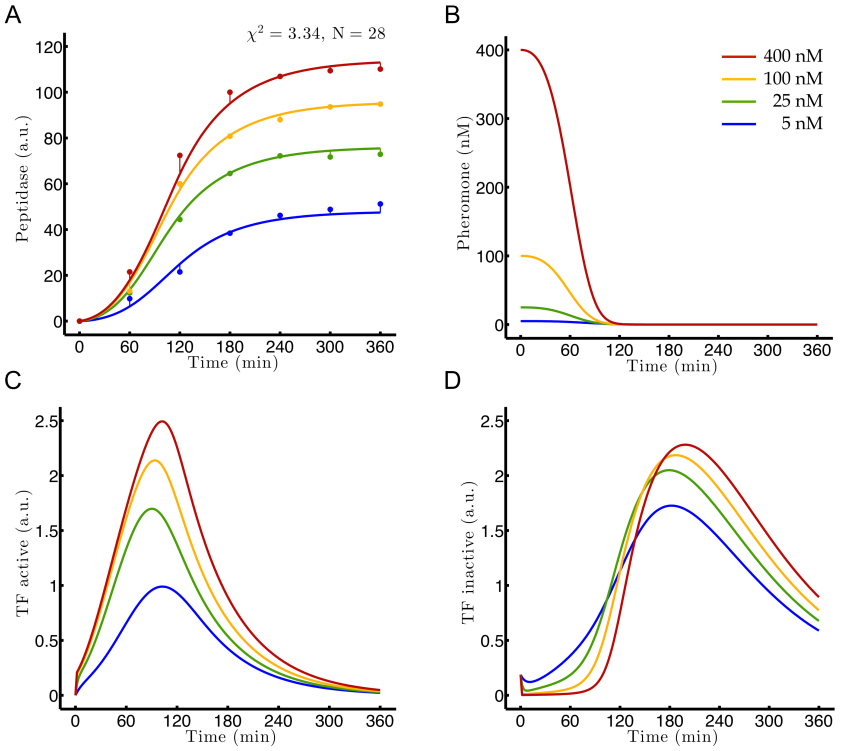


FIGURE 3.11: Simulation of the base model fitted to Ladds *et al.* (1996) Sxa2 data. (A) Trajectories of Peptidase fitted to Sxa2 data. (B-D) Trajectories of unobserved species. (B) Pheromone. (C) TF-active. (D) TF-inactive.

values of cross-correlations between 0 and 1 (Figure 3.12 A), where each element along the principal diagonal contains the correlation coefficient of a parameter with itself $\kappa_{ii} = 1$. Choosing a level of significance alpha, significantly correlated pairs of parameters can be identified by the appropriate two-tailed t-distribution (Figure 3.12 B).

Although model A1B1C1D1E1 is already a simple description of the pheromone signalling feedback structure, it was found that many of the parameters in this model exhibit significant correlations with other parameters during estimation with the Ladds *et al.* (1996) data, implying that they are weakly determinable (Figure 3.12 B). In particular, parameters k_b, v_1, and k_{m1} share a correlation of $\kappa = 1$ between them, which may mean they are non-identifiable. These parameters are all involved in the production of TF-inactive, suggesting that measurements of Sxa2 contain little information about the expression of Ste11.

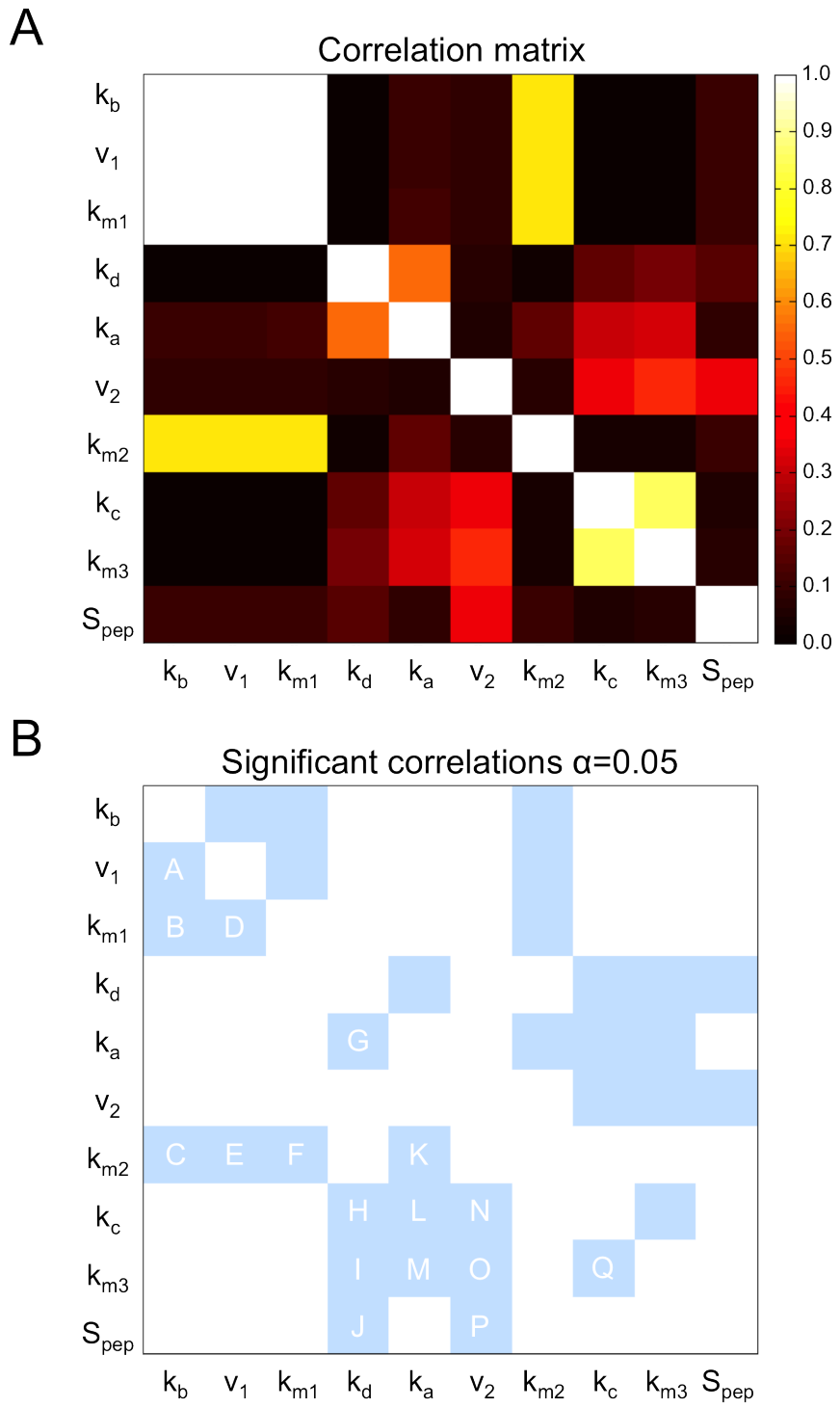


FIGURE 3.12: Correlation between estimated parameters. To assess parameter determinability the cross-correlations between parameters is calculated from the multiple optimization runs performed during parameter estimation. (A) Correlation matrix. (B) The correlation matrix recoloured to show only the significant correlations (blue).

A Monte Carlo-based approach can be used to visualize the functional relationships

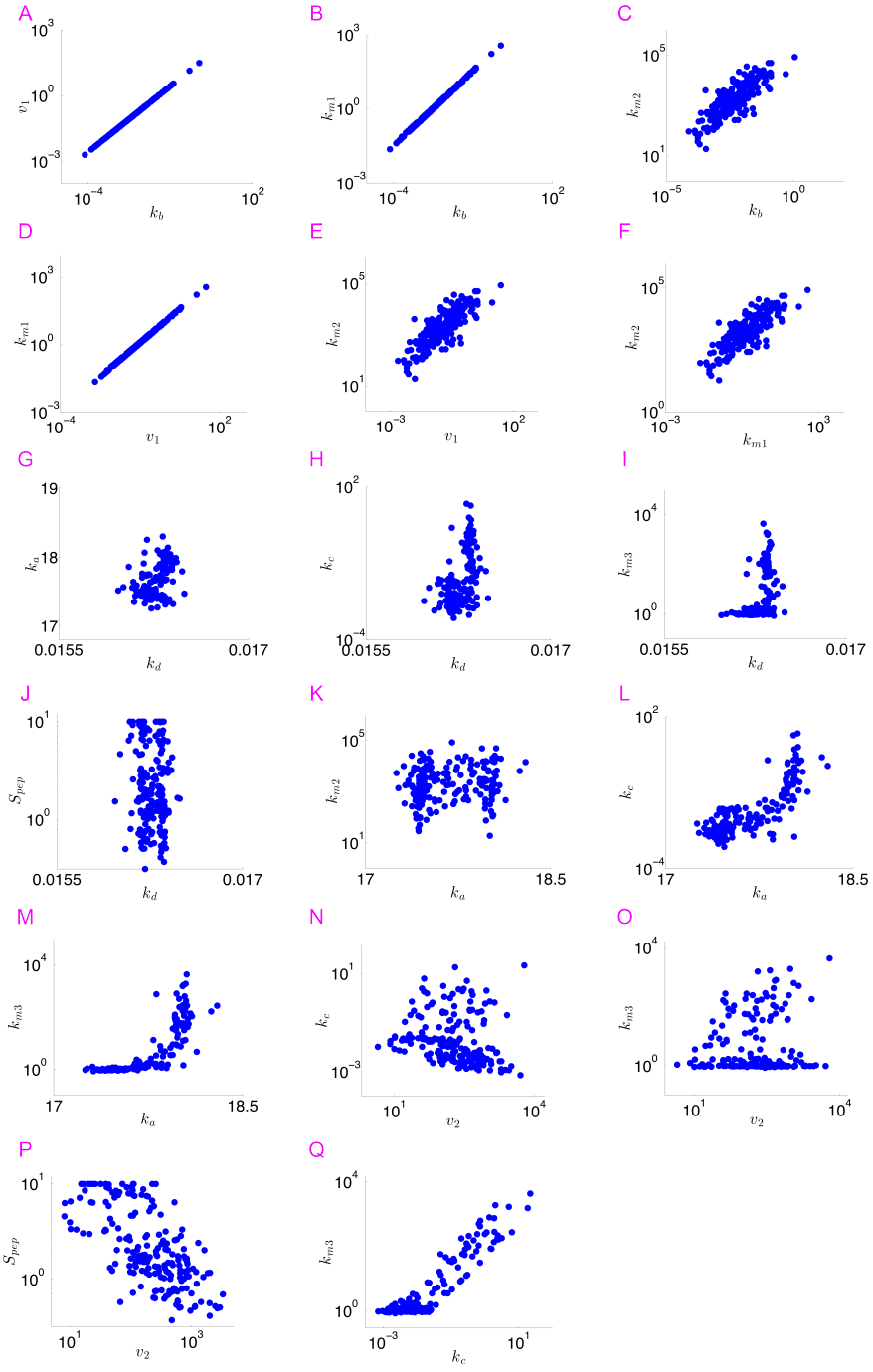


FIGURE 3.13: Monte Carlo analysis of significant correlations from Figure 3.12 B. Each data point represents the final estimated value for the given parameter pair during an individual optimization run.

between the pairs of correlated parameters (Balsa-Canto *et al.*, 2008). By repeating the parameter estimation procedure sufficiently many times (typically several hundreds), the scatter plots generated from estimated values for each correlated pair will display a cloud of solutions whose shape is diagnostic of the relationship

(Balsa-Canto and Banga, 2010). Focusing on the 17 pairs of significantly correlated parameters found previously (A-Q in Figure 3.12 B), 1000 trials of parameter estimation were performed and scatter plots built considering only converged fits (Figure 3.13).

As expected, parameter pairs A, B, and D display a very strong correlation, with all the solutions falling along a straight line with no appreciable bounds. This suggests the existence of a functional relationship between them causing a structural non-identifiability. Correlations involving parameter k_{m2} (C, E, F) are not as severe as the first group, with the solutions contained inside an ellipse, and possibly a lower bound being defined, but the large range of possible values mean that it is not identifiable in practice.

Weakly correlated pairs tend to display a spherical arrangement in their solutions, as in pair G, indicating that the correlation is not severe enough to prevent identifiability. Parameters k_d and k_a only appear in weak correlations with other parameters (G, H, I, J, K, L, M), and are always constrained to a very small interval, which suggests they are identifiable. Parameter k_{m3} appears to have a clearly defined lower bound, but no upper bound (I, M, O, Q). Parameter S_{pep} is by definition structurally non-identifiable in this case, because it is used for scaling of the only data used for estimation, so to limit regression instabilities it was restricted to a smaller range than the rest of the parameters (Jaqaman and Danuser, 2006).

3.4.1 Confidence intervals

To make rigorous statements about the identifiability of the parameters, confidence intervals can be calculated for the estimated parameters based on several metrics. The most common approach is the calculation of asymptotic confidence intervals derived from curvature measures of the likelihood such as the Hessian or Fisher

Information Matrix (Ljung, 1987); however, these measures are not exact when there are non-linear relationships between the observables and the parameters (Schenkendorf *et al.*, 2009; Baker *et al.*, 2015). In addition, it has been noted that asymptotic confidence intervals are not a good approximation of the true variance in estimated parameters when small samples of experimental data are used for parameter estimation or when measurement error is large (Raue *et al.*, 2011), both scenarios being common in biological experimentation.

3.4.2 Profile likelihood estimation

Due to the limitations of standard intervals discussed above, here we use the more appropriate likelihood-based confidence intervals to quantify the uncertainty in the values of estimated parameters, as they are in general more robust against small or noisy samples (Neale and Miller, 1997).

As mentioned in section 3.3, during parameter estimation the difference (residuals) between the experimental data and the model is minimized by non-linear least squares, and when measurement error is normally distributed this problem is equivalent to maximum likelihood estimation, which implies that the residuals will follow a χ^2 distribution (Raue and Timmer, 2011). Likelihood-based confidence intervals can then be generated by defining a threshold Δ_α in the likelihood, that corresponds to the $1-\alpha$ quantile, of the χ^2 distribution (Raue *et al.*, 2009). A confidence region is then defined by

$$\left\{\theta \mid \chi^2(\theta) - \chi^2(\hat{\theta}) < \Delta_\alpha\right\} \quad \text{with} \quad \Delta_\alpha = Q\left(\chi^2, 1-\alpha\right) \tag{3.22}$$

where the borders of this region are the confidence interval $\sigma_i^{\pm}$ with certainty level $1-\alpha$. Although the confidence region extends across multiple dimensions (as many as the number of parameters), the borders will ultimately be determined by the

flattest direction in the likelihood, i.e. the worst constrained. These flat trajectories can be found for each parameter θ_i by profile likelihood estimation (PLE), where the likelihood is explored in the direction of least change by considering all other parameters in the model as nuisance parameters (not of immediate interest but necessary to analyse the parameter of interest):

$$\chi^2_{\mathrm{PL}}(\theta_i) = \min_{\theta_{j\neq i}} \left[\chi^2(\theta) \right] \tag{3.23}$$

To calculate the profile likelihood (χ^2_{PL}) for a proposed optimum, at each step the parameter of interest θ_i is kept fixed, while the nuisance parameters $\theta_{j\neq i}$ are allowed to vary during a re-optimization of the objective function $\chi^2(\theta)$. The process is repeated with step-wise increases in the value of θ_i until the resulting χ^2 value crosses the threshold established by (3.22), or the upper bound of the allowed search space is reached. Then χ^2_{PL} is sampled in the opposite direction of decreasing θ_i.

Concrete statements about identifiability of the parameters can then be made according to the nature of the confidence intervals that are found. Structural non-identifiabilities are manifested as perfectly flat χ^2_{PL} trajectories that are insensitive to parameter value, and that extend to infinity in both directions. This means that at every step of re-optimization, a combination of one or more nuisance parameters was able to fully compensate an increase or decrease in θ_i, leaving χ^2 unchanged. Practical non-identifiabilities are identified by a flattening of the profile likelihood before reaching the threshold Δ_α. Here the χ^2_{PL} trajectories are not completely insensitive to changes in θ_i, and in some cases finite upper or lower bounds may be found, but the confidence interval in one direction of θ_i remains infinite. A fully identifiable parameter then, is characterised by confidence intervals with finite upper and lower bounds.

The profile likelihood was calculated for all parameters in model A1B1C1D1E1 fitted to the Ladds *et al.* (1996) data, and their likelihood-based confidence intervals were assessed to a 95% confidence level (Figure 3.14). Parameters v_1, k_{m1}, k_{m2}, and S_{pep} all have confidence intervals that extend indefinitely in both directions within the limits that were sampled. The χ^2_{PL} of v_1 and S_{pep} is always kept flat, consistent with structural non-identifiabilities, while for k_{m1} and k_{m2} some sensitivity is displayed towards the edges of their allowed range. The search was not continued in those directions because values that are beyond five orders of magnitude with respect to the other parameters in the model will be either too fast or too slow for the time scale considered and could be ignored by further model reduction (Raue *et al.*, 2011).

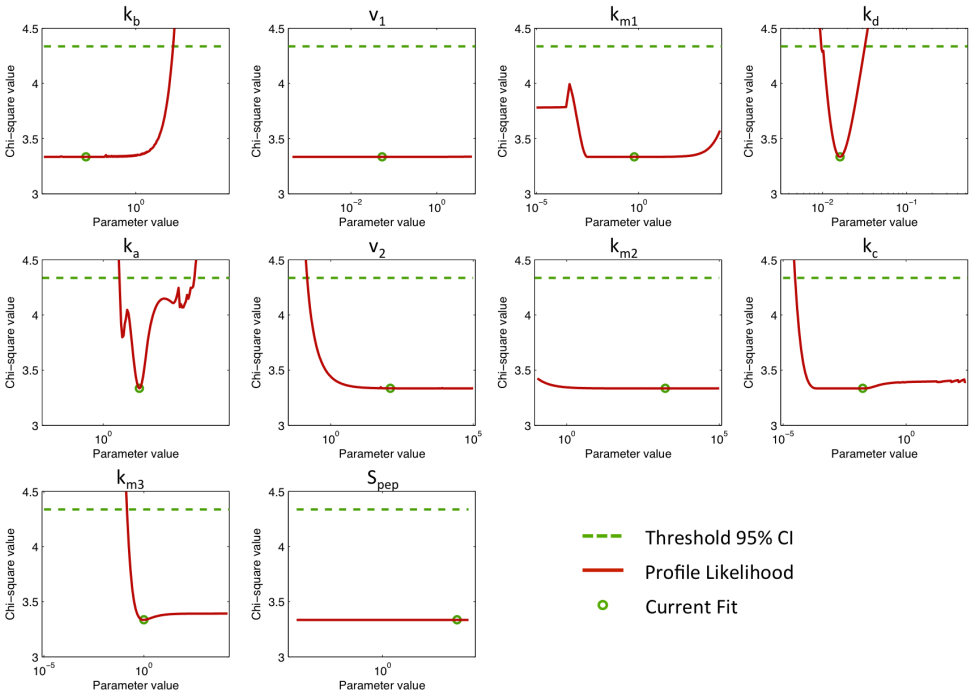


FIGURE 3.14: Profile likelihood estimation for model parameters. Starting from the values obtained during parameter estimation (green circles, Table 3.3), the profile likelihood trajectories (red lines) are calculated independently for each parameter. The point at which the profile likelihood crosses the threshold of predetermined confidence level represents the confidence interval boundary in that direction.

Parameters k_b, v_2, k_c, and k_{m3} have one finite bound and another infinitely extended bound, corresponding to practical non-identifiabilities. While the lower bound of k_{m3} was already apparent from the Monte Carlo analysis (Figure 3.13), the bounds for k_b, v_2, and k_c found by the profile likelihood method were not as easy to distinguish beforehand. Parameters k_d and k_a both have finite upper and lower bounds which are also well constrained, so they are considered identifiable parameters. The confidence intervals determined by PLE are summarised in Table 3.3.

TABLE 3.3: Individual confidence intervals $[\sigma^-, \sigma^+]$ to 95% confidence level determined by PLE for parameters of model A1B1C1D1E1 fitted to Ladds *et al.* (1996) data. Values are given in $\log_{10}$ scale.

Name	$\hat{\theta}_i$	Non-identifiability	σ^-	σ^+
k_b	-2.71	Practical	$-\infty$	$+2.03$
v_1	-1.27	Structural	$-\infty$	$+\infty$
k_{m1}	-0.22	Practical	$-\infty$	$+\infty$
k_d	-1.79	Identifiable	-2.01	-1.50
k_a	$+1.24$	Identifiable	$+0.56$	$+3.13$
v_2	$+2.09$	Practical	-0.83	$+\infty$
k_{m2}	$+3.23$	Structural	$-\infty$	$+\infty$
k_c	-1.76	Practical	-4.51	$+\infty$
k_{m3}	$+0.00$	Practical	-0.85	$+\infty$
S_{pep}	$+0.87$	Structural	$-\infty$	$+\infty$

The results of the PLE confirm that several of the parameters are not identifiable. This uncertainty in model parameters propagates to the trajectories of state variables, and can affect not only the scaling factor of the model but also qualitative model behaviour, thus limiting the usefulness of the model to investigate the dynamics of interest. The consequences of parameter uncertainty can be visualized by simulating the model along the profile likelihood of a given parameter, for

example, the lack of a lower bound in k_b affects the timing at which peak levels of TF-active are reached, the initial behaviour of TF-inactive, the time at which all pheromone has been inactivated, as well as the concentration scale on which events occur (Figure 3.15).

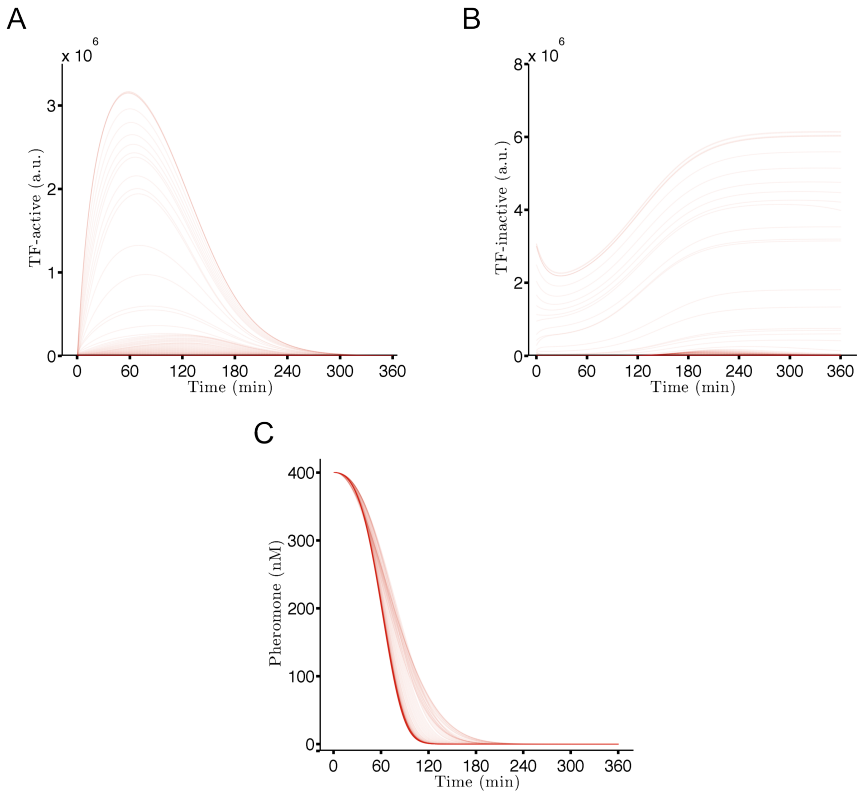


FIGURE 3.15: Consequences of non-identifiability. Model trajectories of unobserved species TF-active (A), TF-inactive (B), and Pheromone (C), are plotted along the profile likelihood of parameter k_b to visualize the propagation of uncertainty from parameters to internal model states. For clarity only the uncertainty for a pheromone dose of 400 nM is shown.

Throughout the initial model assessment carried out in this chapter a recurring observation has been made that the Ladds *et al.* (1996) data set does not sufficiently constrain the model to infer properties of the system of interest. However, before embarking on the generation of additional data sets to supplement parameter estimation, the finding of structural identifiabilities in the model raises the question of whether the parameters will be identifiable, even in principle, from the set of measurements that will be chosen. It is possible for the model to remain unidentifiable even when all state variables are directly measurable (Quaiser and

Mönnigmann, 2009). In practice the choice of measurable species is limited, and often depends on external factors such as availability of necessary technologies and budget considerations.

To answer this question in an exact way requires a structural identifiability analysis (Cobelli and DiStefano III, 1980). Practical identifiability analysis is necessarily a local method, and its results can only be trusted up to a certain level of confidence (Raue *et al.*, 2014). Structural identifiability results hold for the global parameter space and are exact, but they also involve a significant increase in calculation complexity that very quickly becomes intractable even for models of modest size (Chis *et al.*, 2011). To inform the selection of measurements and the possibilities of obtaining a fully identifiable model, a structural identifiability analysis is described in the next section.

3.5 Structural identifiability analysis

There are several well-established techniques for performing structural identifiability analysis of linear models, however, most models in systems biology are nonlinear. The choice of technique for nonlinear models is often very limited, due to restrictive assumptions about the model structure or intractable computational complexity. To analyse the model presented in Equations (3.1)-(3.5), we employed the similarity transform approach, which is suitable for autonomous nonlinear rational systems (Evans *et al.*, 2002).

3.5.1 Similarity transformation approach for uncontrolled systems (STAUS)

This technique of structural identifiability analysis considers models of the following form:

$$\dot{\boldsymbol{x}}(t,\boldsymbol{p}) = \boldsymbol{f}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}) \tag{3.24}$$

$$\boldsymbol{x}(0,\boldsymbol{p}) = \boldsymbol{x_0}(\boldsymbol{p}) \tag{3.25}$$

$$\boldsymbol{y}(t,\boldsymbol{p}) = \boldsymbol{h}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}) \tag{3.26}$$

where $\boldsymbol{p} \in \mathbb{R}^q \subset \boldsymbol{\Omega}$ is the vector of unknown parameters, $\boldsymbol{x}(t,\boldsymbol{p}) \in \mathbb{R}^n$ is the state vector with initial conditions $\boldsymbol{x_0}(\boldsymbol{p})$, and $\boldsymbol{y}(t,\boldsymbol{p}) \in \mathbb{R}^r$ is the vector of observables. It is assumed that $\boldsymbol{f}(\,\cdot\,,\cdot\,)$ and $\boldsymbol{h}(\,\cdot\,,\cdot\,)$ are rational functions in both $\boldsymbol{x}$ and $\boldsymbol{p}$, while $\boldsymbol{x_0}$ is a rational function in $\boldsymbol{p}$.

In the remainder of the analysis we use the following definitions from Evans *et al.* (2002):

Definition 1. Parameter vectors $\boldsymbol{p}$, $\bar{\boldsymbol{p}} \in \boldsymbol{\Omega}$ are said to be indistinguishable, written $\boldsymbol{p} \sim \bar{\boldsymbol{p}}$, if there exists a $\tau > 0$ such that $\boldsymbol{y}(t,\boldsymbol{p}) = \boldsymbol{y}(t,\bar{\boldsymbol{p}})$ for all $t \in [\,0, t\,)$.

Definition 2. A model of the form (3.24)-(3.26) is said to be globally identifiable at $\boldsymbol{p} \in \boldsymbol{\Omega}$ if $\bar{\boldsymbol{p}} \in \boldsymbol{\Omega}$ and $\boldsymbol{p} \sim \bar{\boldsymbol{p}}$ imply that $\boldsymbol{p} = \bar{\boldsymbol{p}}$. If this is only true in some neighbourhood of $\boldsymbol{p}$ then the model is locally identifiable at $\boldsymbol{p} \in \boldsymbol{\Omega}$.

Definition 3. If (3.24)-(3.26) is globally (locally) identifiable at generic $\boldsymbol{p}$ for almost all $\boldsymbol{p} \in \boldsymbol{\Omega}$, then it is said to be structurally globally (locally) identifiable. The model is said to be unidentifiable if it is not structurally locally identifiable.

The analysis proceeds as follows, first the model is checked to satisfy a prerequisite observability condition. If this condition satisfied, then the existence of a smooth mapping between the state trajectories generated by indistinguishable parameter

vectors is guaranteed. Then, based on this mapping, a series of equations can be derived from the initial conditions, model structure, and observation function, that relate any given pair of indistinguishable parameter vectors. Finally, by solving these equations simultaneously, the resulting relationship between indistinguishable parameter vectors determines the identifiability of the model.

3.5.2 The Observability Rank Condition (ORC)

For a generic parameter vector $\boldsymbol{p}$, the following function is defined:

$$\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{p}) = \big(\mu_1(\boldsymbol{x}, \boldsymbol{p}), \ldots, \mu_n(\boldsymbol{x}, \boldsymbol{p})\big)^{\mathsf{T}} \tag{3.27}$$

where the first r smooth functions are given by $\mu_i(\boldsymbol{x}, \boldsymbol{p}) = h_i(\boldsymbol{x}, \boldsymbol{p})$ (for r number of outputs), and the remaining $n - r$ functions (for n number of states) can be chosen from the infinite list of functions ϕ_i obtained from the Lie derivatives of the μ_i

$$\phi_i(\boldsymbol{x}, \boldsymbol{p}) = \mathcal{L}_f \mu_i(\boldsymbol{x}, \boldsymbol{p}) = \frac{\partial \mu_i}{\partial \boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{p}) \cdot \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{p}) \tag{3.28}$$

for $i = 1, \ldots, r$, and then recursively by

$$\phi_i(\boldsymbol{x}, \boldsymbol{p}) = \mathcal{L}_f \phi_{i-r}(\boldsymbol{x}, \boldsymbol{p}) \tag{3.29}$$

for $i > r$, and where the elements of $\boldsymbol{f}$ are the system coordinate functions given by (3.24). If the Jacobian matrix of $\boldsymbol{H}$ with respect to $\boldsymbol{x}$, evaluated at $\boldsymbol{x_0}(\boldsymbol{p})$ is non-singular, then the system (3.24)-(3.26) is said to satisfy the Observability Rank Condition (ORC).

In the case of model (3.1)-(3.5), no output structure was defined, but we can test different combinations of observables based on the measurement technologies available to us. In the first instance we assume all species are directly observable

except for the differentially active species of Ste11, which can only be detected by measuring the total amount of Ste11 (e.g. Western blot):

$$\boldsymbol{y} = \begin{bmatrix} x_1 + x_2 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} = \boldsymbol{H} \tag{3.30}$$

The Jacobian matrix of $\boldsymbol{H}$ with respect to $\boldsymbol{x}$ is given by

$$\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial \mu_1}{\partial x_1} & \frac{\partial \mu_1}{\partial x_2} & \frac{\partial \mu_1}{\partial x_3} & \frac{\partial \mu_1}{\partial x_4} \\ \frac{\partial \mu_2}{\partial x_1} & \frac{\partial \mu_2}{\partial x_2} & \frac{\partial \mu_2}{\partial x_3} & \frac{\partial \mu_2}{\partial x_4} \\ \frac{\partial \mu_3}{\partial x_1} & \frac{\partial \mu_3}{\partial x_2} & \frac{\partial \mu_3}{\partial x_3} & \frac{\partial \mu_3}{\partial x_4} \\ \frac{\partial \mu_4}{\partial x_1} & \frac{\partial \mu_4}{\partial x_2} & \frac{\partial \mu_4}{\partial x_3} & \frac{\partial \mu_4}{\partial x_4} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \boldsymbol{J} \tag{3.31}$$

Evaluating at the initial condition (3.5) leaves $\boldsymbol{J}$ unchanged, and having full rank it satisfies the ORC. Assuming an observation gain for all measurements has no effect on this result, however, we find that all state variables must be measured, at least in some combination (e.g. as in μ_1) in order satisfy this condition.

Having satisfied the ORC, we proceed with the analysis by utilising theorem 6 from Evans *et al.* (2002):

Theorem 6. *For $\boldsymbol{p} \in \boldsymbol{\Omega}$, let $\mu_1, \ldots, \mu_n$ be smooth functions for which* (3.24)-(3.26) *satisfies the ORC at $\boldsymbol{x_0}(\boldsymbol{p})$, and $\boldsymbol{H}$ the corresponding function defined in* (3.27)*. If $\bar{\boldsymbol{p}} \in \boldsymbol{\Omega}$, then $\boldsymbol{p} \sim \bar{\boldsymbol{p}}$ if and only if there exists a neighbourhood $\boldsymbol{V}_{\bar{\boldsymbol{p}}}$ of $\boldsymbol{x_0}(\bar{\boldsymbol{p}})$, a $\tau > 0$, and a C^∞ map $\lambda : \boldsymbol{V}_{\bar{\boldsymbol{p}}} \mapsto M(\boldsymbol{p})$* such that

$$\boldsymbol{H}_{\boldsymbol{p}}\big(\lambda(\boldsymbol{x})\big) = \boldsymbol{H}_{\bar{\boldsymbol{p}}}(\boldsymbol{x}) \tag{3.32}$$

for all $\boldsymbol{x} \in \boldsymbol{V}_{\bar{\boldsymbol{p}}}$*, and*

$$\lambda\big(\boldsymbol{x_0}(\bar{\boldsymbol{p}})\big) = \boldsymbol{x_0}(\boldsymbol{p}) \tag{3.33}$$

$$\boldsymbol{f}\big(\lambda\big(\boldsymbol{x}(t,\bar{\boldsymbol{p}})\big),\boldsymbol{p}\big) = \frac{\partial\lambda}{\partial\boldsymbol{x}}\big(\boldsymbol{x}(t,\bar{\boldsymbol{p}})\big)\cdot\boldsymbol{f}\big(\boldsymbol{x}(t,\bar{\boldsymbol{p}}),\bar{\boldsymbol{p}}\big) \tag{3.34}$$

$$\boldsymbol{h}\big(\lambda\big(\boldsymbol{x}(t,\bar{\boldsymbol{p}})\big),\boldsymbol{p}\big) = \boldsymbol{h}\big(\boldsymbol{x}(t,\bar{\boldsymbol{p}}),\bar{\boldsymbol{p}}\big) \tag{3.35}$$

for all $t \in [\,0,t\,)$*, where* $\boldsymbol{x}(t,\bar{\boldsymbol{p}})$ *is the solution of system* (3.24)-(3.26) *for parameter vector* $\bar{\boldsymbol{p}}$*.*

For the model (3.1)-(3.5) with observation function (3.30), equation (3.32) yields

$$\boldsymbol{H_p}\big(\lambda(\boldsymbol{x})\big) = \begin{bmatrix} \lambda_1+\lambda_2 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = \begin{bmatrix} x_1+x_2 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \tag{3.36}$$

which gives $\lambda_i(x) = x_i$. Substituting into (3.34) yields for the left hand side

$$\boldsymbol{f}\big(\lambda\big(\boldsymbol{x}(t,\bar{\boldsymbol{p}})\big),\boldsymbol{p}\big) =$$

$$\begin{bmatrix} \dfrac{k_b k_{m_1} - k_{m_1}k_d x_1 + k_b x_2 + v_1 x_2 - k_d x_1 x_2 - k_{m1}k_a x_1 x_4 - k_a x_1 x_2 x_4}{k_{m_1}+x_2} \\ k_a x_1 x_4 - k_d x_2 \\ \dfrac{v_2 x_2}{k_{m_2}+x_2} \\ -\dfrac{k_c x_3 x_4}{k_{m_3}+x_4} \end{bmatrix} \tag{3.37}$$

and for the right hand side

$$\frac{\partial \lambda}{\partial \boldsymbol{x}}\big(\boldsymbol{x}(t,\bar{\boldsymbol{p}})\big)\cdot \boldsymbol{f}\big(\boldsymbol{x}(t,\bar{\boldsymbol{p}}),\bar{\boldsymbol{p}}\big) =$$

$$\begin{bmatrix} \dfrac{\bar{k}_b\bar{k}_{m_1} - \bar{k}_{m_1}\bar{k}_d x_1 + \bar{k}_b x_2 + \bar{v}_1 x_2 - \bar{k}_{d_1} x_1 x_2 - \bar{k}_{m1}\bar{k}_a x_1 x_4 - \bar{k}_a x_1 x_2 x_4}{\bar{k}_{m_1} + x_2} \\ \\ \bar{k}_a x_1 x_4 - \bar{k}_d x_2 \\ \\ \dfrac{\bar{v}_2 x_2}{\bar{k}_{m_2} + x_2} \\ \\ -\dfrac{\bar{k}_c x_3 x_4}{\bar{k}_{m_3} + x_4} \end{bmatrix} \tag{3.38}$$

Equating the monomials in x, and solving simultaneously for the parameters yields $k_b = \bar{k}_b$, $v_1 = \bar{v}_1$, $k_{m_1} = \bar{k}_{m_1}$, $k_d = \bar{k}_d$, $k_a = \bar{k}_a$, $v_2 = \bar{v}_2$, $k_{m_2} = \bar{k}_{m_2}$, $k_c = \bar{k}_c$, and $k_{m_3} = \bar{k}_{m_3}$. Thus, the model (3.1)-(3.5) is structurally globally identifiable with the observables (3.30). The analysis was repeated using different combinations of species in the observation function (Table 3.4); however, in all cases it was necessary to include all species, at least in some combination (e.g. $x_1 + x_2$), for the model to be structurally globally identifiable.

TABLE 3.4: Summary of STAUS identifiability analysis for model A1B1C1D1E1. SGI: Structurally globally identifiable. SU: Structurally unidentifiable.

Observation	Result
$y = [\, x_1 \,,\, x_2 \,,\, x_3 \,,\, x_4 \,]^{\mathsf{T}}$	SGI
$y = [\, x_1 + x_2 \,,\, x_2 \,,\, x_3 \,,\, x_4 \,]^{\mathsf{T}}$	SGI
$y = [\, x_1 + x_2 \,,\, x_3 \,,\, x_4 \,]^{\mathsf{T}}$	SGI
$y = [\, x_1 + x_2 \,,\, x_3 \,]^{\mathsf{T}}$	SU
$y = [\, x_1 + x_2 \,,\, x_4 \,]^{\mathsf{T}}$	SU
$y = [\, x_1 \,,\, x_3 \,,\, x_4 \,]^{\mathsf{T}}$	SU
$y = [\, x_2 \,,\, x_3 \,,\, x_4 \,]^{\mathsf{T}}$	SU
$y = [\, x_3 \,,\, x_4 \,]^{\mathsf{T}}$	SU

Chapter 4

Generating data sets for parameter estimation

4.1 Development of an assay to track pheromone depletion

The results of the structural identifiability analysis highlighted the need to obtain measurements for all species in the model to allow the reliable estimation of model parameters (section 3.5). Since Ladds *et al.* (1996) provided information about the production of Sxa2 in response to P-factor, it was desirable to complement this data with the matching decrease in pheromone concentration.

From the outset it was identified that tracking changes in P-factor concentration due to Sxa2 activity requires addressing two main problems. First, at each measured time point the enzymatic activity of Sxa2 must be quickly and irreversibly inactivated to preserve the amount of pheromone present at that moment, and the inactivation method should not affect the integrity or concentration of pheromone in any way. Secondly, the method used to measure the remaining concentration

of pheromone in the samples must be able to discriminate between active and inactive P-factor, which only differ by one amino acid (Ladds *et al.*, 1996).

The first problem was solved by noting that Sxa2 can be efficiently inactivated by heat treatment above 80°C (Ladds *et al.*, 1996). Although it takes 15 min to make the modification irreversible, Sxa2 is unable to continue cleaving P-factor at this temperature, so that pheromone concentration is immediately fixed upon reaching 80°C. Furthermore, since P-factor is a 23 amino acid peptide with no significant secondary structure (Ladds, 1998), it is highly resistant to heat and retains full activity even if autoclaved (Ladds, 1998), which means that Sxa2 heat inactivation does not affect the concentration of active P-factor.

For the second problem, it would be possible in principle to use a reverse-phase high performance liquid chromatography approach, however, since the necessary equipment was not readily available, it was decided to exploit the fact that yeast cells themselves are excellent detectors of active P-factor. A number of reporter strains have been described that can provide dose-response information (Didmon *et al.*, 2002; Das *et al.*, 2006; Smith, 2009; Croft, 2012); however, the simplest option is provided by the fact that strains without a functional *sxa2* gene exhibit a reproducible increase in cell volume when stimulated with mating pheromones, and the cell volume increase exhibits a dose-response relationship with the concentration of pheromone in the medium (Davey, 1991; Ladds *et al.*, 1996).

Based on these observations, an assay to track the depletion of a pheromone input over time was devised (Figure 4.1). The assay begins by adding a particular dose of pheromone to a culture of the test strain, which contains a wild type allele of the Sxa2 peptidase. After P-factor stimulation, samples are removed at the desired time points, and processed immediately after collection by removing cells through filtration or centrifugation, followed by heat treatment to inactivate Sxa2. Removing the cells from the test culture is necessary to ensure that no residual

Sxa2 activity will be carried on to further steps. To ensure that heat transfer is fast and evenly distributed, heat inactivation should be performed on a water bath, and not on a dry heat block.

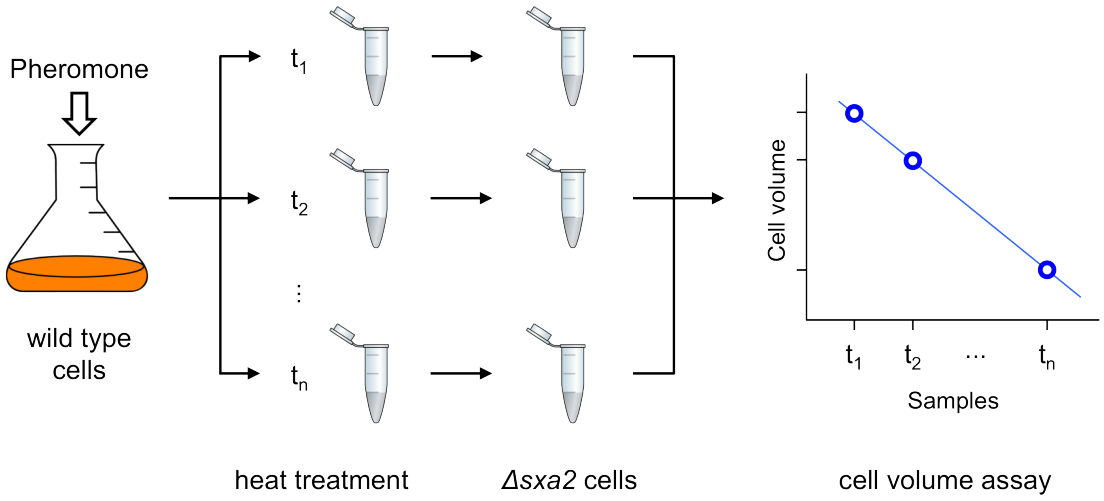


FIGURE 4.1: Experimental design to track P-factor pheromone depletion by Sxa2. After stimulating the test culture with pheromone, samples are collected at time intervals and immediately heat treated to inactivate Sxa2. To determine the amount of pheromone that remains in each sample, the media is transferred to new cultures of *sxa2*$^-$ cells and the cell volume response is measured.

Since P-factor is a very stable peptide, samples can remain at room temperature until all time points have been collected. In parallel to the test culture, a measuring strain which contains a non-functional allele of Sxa2 is cultured to a density of $\sim 10^6$ cells/mL, and is used to inoculate the test media which will contain varying amounts of remaining P-factor. The cells are then incubated under normal growth conditions for 16 hrs to allow a full morphological response to occur, and the median cell volume of each sample is recorded using a cell counter.

To convert between cell volumes of the measuring strain and the concentration of pheromone that was remaining in the test samples, a standard curve should be prepared at time 0 of the experiment under identical conditions to the test samples, and an appropriate dose-response curve is fitted to the standard curve to interpolate pheromone concentrations from cell volumes.

Because experiments with *sxa2*$^{+}$ strains are not common in pheromone signalling research, a number of conditions had to be optimised for this assay which are normally not relevant.

4.1.1 Minimising pheromone loss due to adsorption

In standard experiments, pheromone stocks are prepared in methanol and the required dose is delivered by transferring the necessary volumes of diluted stock into empty tubes, followed by air drying to remove the methanol vehicle. Test cultures are then added to the tubes for incubation with the pheromone. With this method of delivery, all the pheromone is adsorbed on to the tube walls, and the concentration of pheromone will not be uniform as perceived by the cells in the liquid medium. This is usually not a problem, since these experiments are only concerned with steady state responses, and any inhomogeneities are averaged out over the long periods of incubation used, which are typically around 16 hrs long.

For the purposes of this work, the interest is on the transient behaviour that begins immediately after pheromone stimulation, which makes the traditional method of dose delivery inadequate. To allow an even concentration of P-factor to be perceived by all cells in the culture, the pheromone dose must be delivered in solution. This also ensures that Sxa2 will have unrestricted access to all P-factor molecules, which may otherwise be protected from cleavage. However, since P-factor is highly hydrophobic (Imai and Yamamoto, 1994), it has a natural tendency of dropping out of solution by adsorbing to the walls of the container, which would again affect the concentration of pheromone in solution, and add variability to the results since the assay requires transferring the test media to a new tube at least once.

The impact of adsorption on P-factor concentration was assessed by preparing tubes containing DMM minimal media and adding P-factor dissolved in methanol

to a final concentration of 1 µM. The tubes were then incubated in standard yeast growth conditions, and every 30 min the contents of one sample were transferred to a new tube, for a total of 4 hrs. After collecting all samples, the media was inoculated with a reporter strain that expresses *lacZ* from the *sxa2* locus (JY544) to measure differences in pheromone between the samples (Figure 4.2).

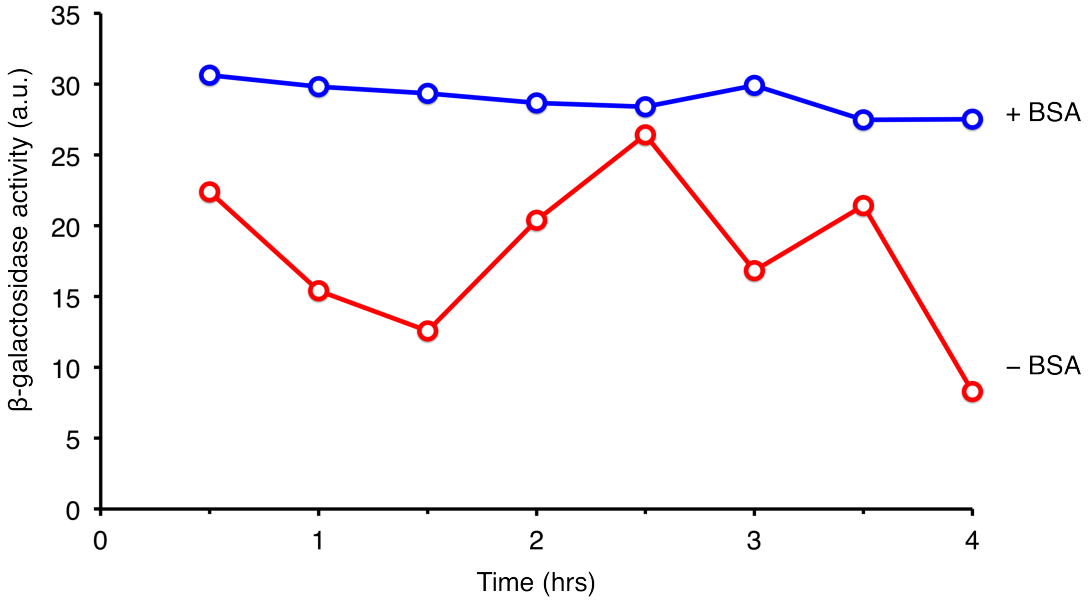


FIGURE 4.2: Pre-coating tubes with BSA minimises pheromone loss due to adsorption. 1 mL aliquots of 1 µM P-factor were incubated at 30°C in standard 1.5 mL microcentrifuge tubes, with or without BSA pre-coating, and at selected time points one of the samples was transferred to a new tube. The concentration of pheromone that was successfully transferred was compared between samples by measuring the response of the JY544 (*sxa2*>*lacZ*) reporter strain.

Pheromone loss due to adsorption was shown to cause a high amount of variability in the concentration of P-factor that is recovered after transferring media between tubes, although pheromone loss was not always proportional to the incubation time (Figure 4.2). To minimise this effect, two different solutions were tested for effectiveness using the same experiment. The first option was the specialised Protein LoBind tubes (Eppendorf, UK), which are manufactured to improve the recovery rate of proteins in solution due to adsorption. The second option was to saturate standard 1.5 mL microcentrifuge tubes with a coating of purified protein to block all plastic surfaces where P-factor could adsorb. For this approach >99%

pure BSA (Sigma-Aldrich, UK) was used dissolved in PBS at 2 mg/mL, and 1 mL of blocking solution was incubated in each tube for 16 hrs in standard yeast growth conditions. Before use in experiments, the blocking solution was removed by aspiration and the tube was allowed to air-dry in a sterile environment.

It as found that BSA coated tubes were the only option able to significantly reduce pheromone loss due to adsorption, with the resulting variability being indistinguishable from the expected variance of the beta-galactosidase assay (Figure 4.2). Under these conditions, Protein LoBind tubes had the same performance as standard tubes without BSA coating.

4.1.2 Effect of methanol concentration on pheromone response

A secondary issue that arose from delivering P-factor in solution directly to cultures was the unexpected side effect of exposing yeast to methanol. Due to its hydrophobicity, lyophilised P-factor cannot be dissolved directly into water and requires an organic solvent to enter into solution (Imai and Yamamoto, 1994). Methanol has proved to be an effective solvent for P-factor; however, it is also employed in many yeast protocols as a fixative agent (Alfa *et al.*, 1993).

It was expected that the standard volumes that are routinely used to dispense P-factor into tubes would be small enough to safely ignore its effects on yeast cells. For dose-response experiments, serial dilutions of P-factor in methanol are usually prepared so that whole-log concentrations require 34 µL per mL of culture, while half-log concentrations require 10 µL per mL of culture; which correspond to proportions of methanol to media of ∼1:30 and 1:100 respectively.

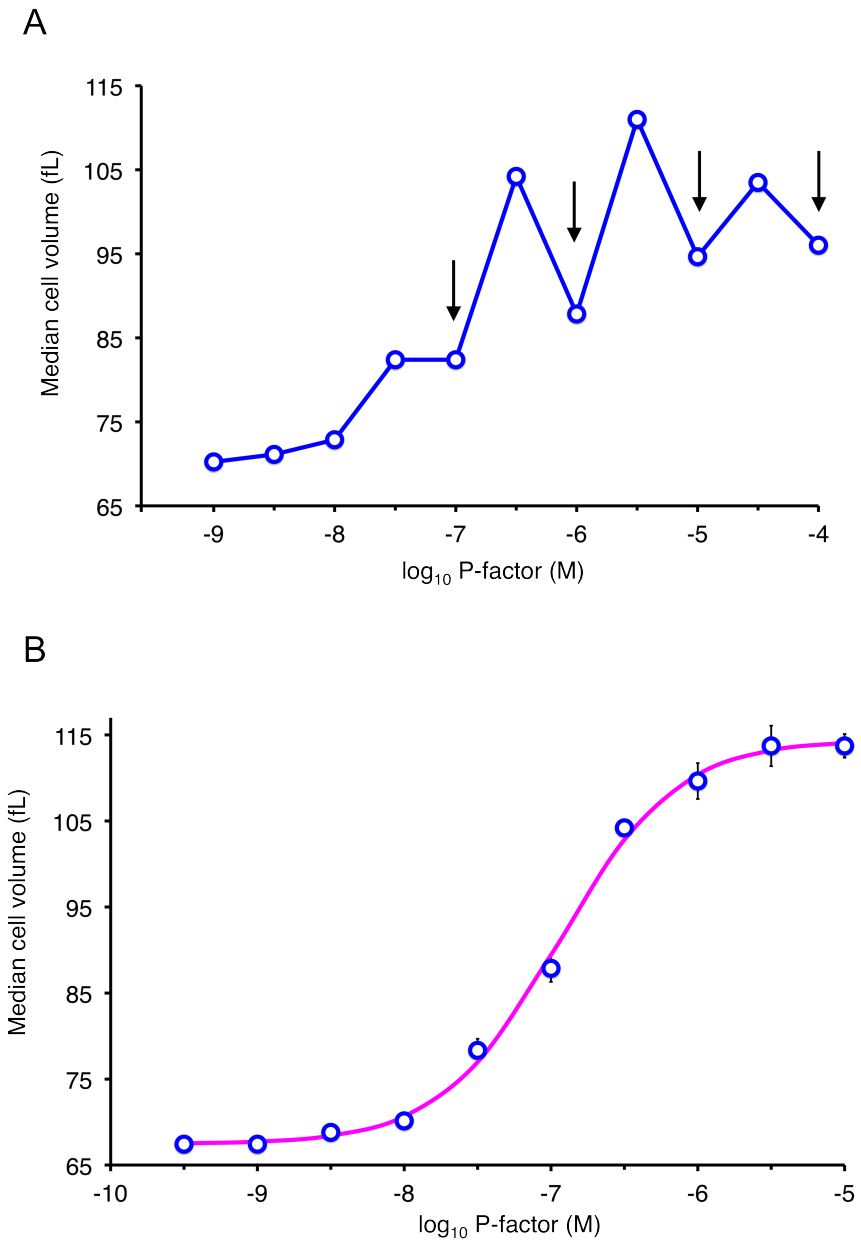


FIGURE 4.3: Methanol affects the ability of cells to respond to pheromone. *sxa2*$^{-}$ cells were stimulated with a range of pheromone concentrations by adding P-factor dissolved in methanol directly to 1 mL culture samples, and the cell volume response was recorded after 16 hrs. (A) To achieve the desired whole-log or half-log final concentrations, pheromone was delivered in 34 or 10 µL of methanol of an appropriate stock serial dilution, however, samples receiving the higher amount of methanol (arrows) displayed a marked decrease in their response. (B) Pheromone delivered in 3.4 µL or 1.0 µL of methanol produced a smooth dose-response to which a sigmoidal curve can be fit.

Performing dose-response experiments with the cell volume assay showed that when using the standard serial dilution volumes for P-factor dosage, cells exposed to 1:30 methanol displayed a marked decrease in their ability to respond to

pheromone (Figure 4.3 A). This adverse effect disappeared for greater dilutions of methanol (Figure 4.3 B), so dosage volumes were changed to employ stock dilutions that were ten-fold more concentrated, to dispense 3.4 and 1.0 µL per mL of culture respectively.

4.1.3 Time course quantification of P-factor depletion by Sxa2

To determine the time frame over which the final assays should be performed, pilot time courses were performed sampling the same time points as the Ladds *et al.* (1996) data set (Figure 4.4). It was found that the greatest power of resolution for this assay occurred between 1 to 2 hrs after pheromone stimulation, where the biggest changes in cell volume were detected (Figure 4.4). Only minor changes were detected during the first hour of the experiment, so it was decided to focus sampling on the period from 60 to 180 min post-stimulation. Reducing the time range of the experiment allowed a greater number of technical replicates to be handled in each assay, and to sample more densely this region of interest.

The sensitivity of the assay allowed two doses of pheromone to be tested reliably, 1 µM and 0.1 µM of P-factor (Figure 4.5). It was found that the dynamics of pheromone clearance end much earlier than Sxa2 production, which still occurs at a smaller rate up to 2-3 hrs after no active pheromone can be detected (Figure 4.5). The initial amount of pheromone also determines the speed of clearance, which for the 1 µM dose 90% of pheromone was inactivated by the 135 min mark, while for the 0.1 µM dose the equivalent proportion was reached 15 min earlier at 120 min (Figure 4.5). For doses below 0.1 µM only step decreases in cell volume could be measured, which are not very informative and so were not pursued further. Doses above 1 µM were not assayed since this dose is able to elicit a maximal

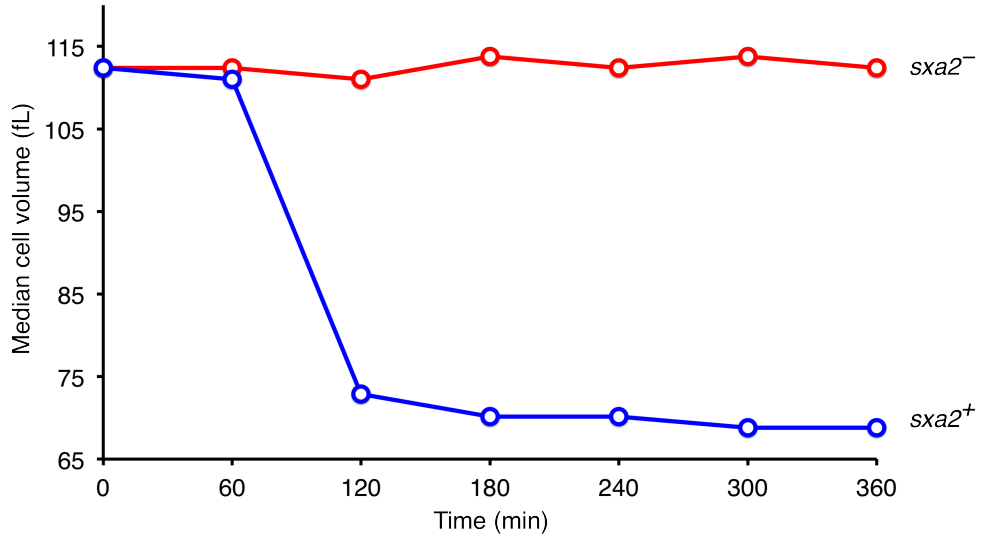


FIGURE 4.4: Most of the pheromone dose inactivation occurs between 1 and 2 hours after stimulation. To determine the most relevant time range to sample for pheromone depletion tracking, a preliminary assay was performed following the Ladds *et al.* (1996) data set, with hourly time points over the course of 6 hrs. To control for non-specific pheromone loss a *sxa2⁻* strain was assayed simultaneously.

transcriptional response, and higher concentrations of pheromone can result in toxic effects (Weston *et al.*, 2013).

4.2 Generating absolute quantification measurements of Sxa2

To fit the relative Ladds *et al.* (1996) data in chapter 3, a scaling parameter S_{pep} had to be introduced to relate the data to the Peptidase state variable. Since by definition relative data contains no information about the scaling parameters, S_{pep} is structurally non-identifiable, and as shown in Figures 3.12 and 3.13, this produces functional relationships with other parameters, which in turn also affects their identifiability.

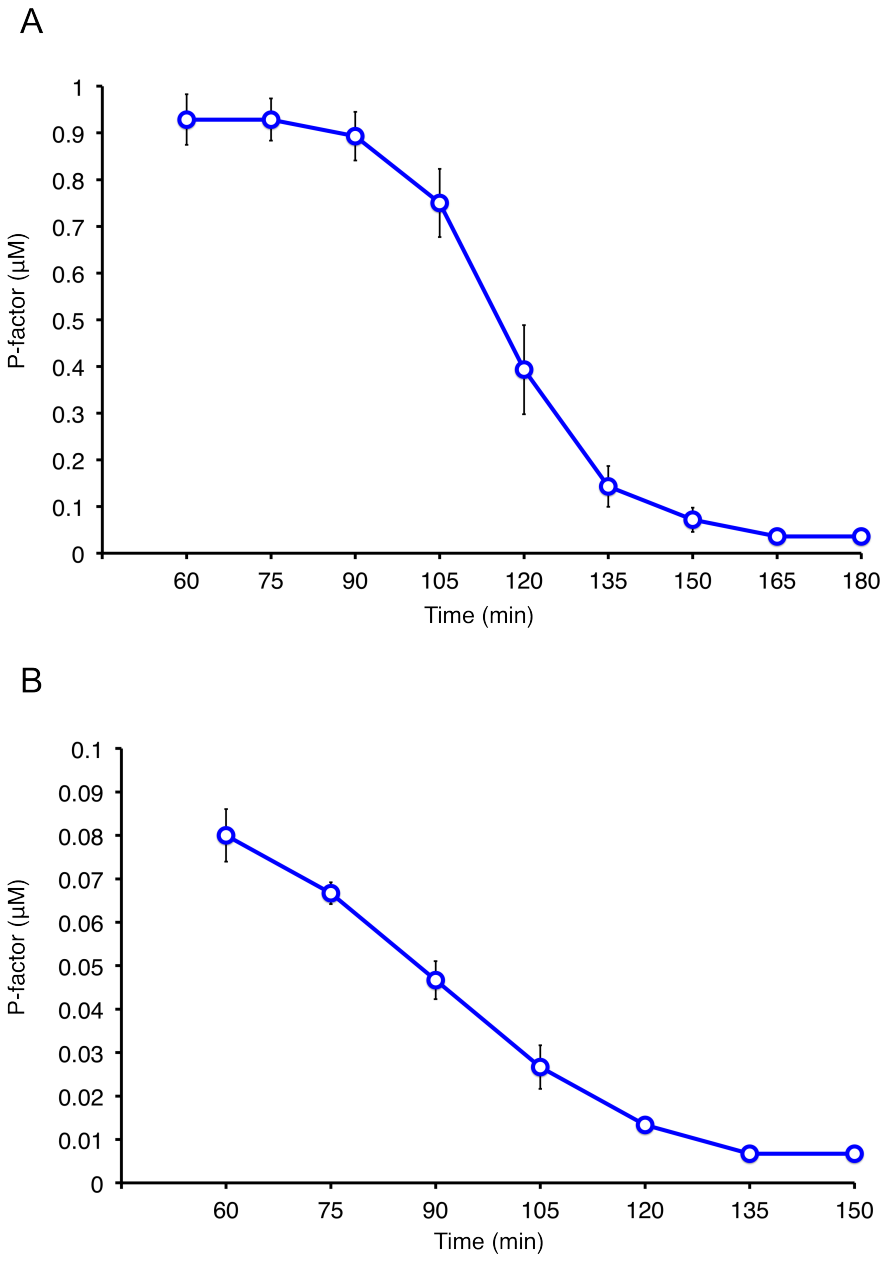


FIGURE 4.5: Absolute quantification time course of P-factor inactivation by Sxa2. Wild type cells (JY486) were cultured in DMM minimal media to mid-log phase ($\sim 5 \times 10^6$ cells/mL) and stimulated with P-factor to a final concentration of either 1 µM (A), or 0.1 µM (B). The pheromone depletion tracking assay was performed on samples collected every 15 min, starting 1 hr after stimulation. Cell volumes were converted to P-factor concentration by interpolation from a sigmoidal dose-response curve fitted to a standard curve constructed at the beginning of each assay. Results shown are means ± SD of three biological replicates.

The only way to resolve the non-identifiability of S_{pep} is to provide the model with absolute quantification measurements of Sxa2 so that the value of S_{pep} can

be estimated by simultaneous fitting to both relative and absolute data. Immunodetection of Sxa2 has been attempted before and has proved to be challenging (Ladds, 1998; Ladds and Davey, 2000). Since Sxa2 is a secreted protein it can be detected directly from the extracellular space without the need to prepare cell extracts, however, the low abundance of secreted Sxa2 compared to the volume of the growth medium makes the concentration of Sxa2 lower than the limits of detection of standard methods (Ladds, 1998).

Successful detection of Sxa2 in the past was achieved by two methods, either by using an inducible promoter to drive ectopic expression of *sxa2* up to ~ 260 times the level of normal expression, or alternatively, by using ultrafiltration membranes to concentrate normal amounts of Sxa2 in media into a smaller volume (Ladds, 1998; Ladds and Davey, 2000). There have been no antibodies developed to target Sxa2, so in both cases detection was made by adding a polyhistidine tag to Sxa2, and using an anti-His antibody for the analysis.

Since the aim was to measure endogenous levels of Sxa2 production in response to pheromone, the overexpression approach was not suitable for this purpose, and further efforts were directed towards the use of ultrafiltration concentrators. To improve the odds of obtaining reliable measurements, commercial solutions for quantitative protein detection were sought to find the most sensitive assays available. The best option found was an anti-GFP ELISA kit (ab171581, abcam, UK), with a limit of detection of 1.8 pg/mL, which is at least three orders of magnitude more sensitive than typical western immunoblotting assays (Hnasko and Hnasko, 2015).

4.2.1 Creating a *sxa2-GFP* yeast strain

To quantify Sxa2 using the anti-GFP ELISA kit required the construction of a *sxa2-GFP* allele, and the creation of a yeast strain expressing *sxa2-GFP* from the

sxa2 locus (Figure 4.6). The Sxa2-GFP fusion protein was designed to have the GFP tag at the C-terminus, since the Sxa2 pre-pro-protein contains a N-terminal signal peptide necessary for secretion (Nielsen *et al.*, 1997).

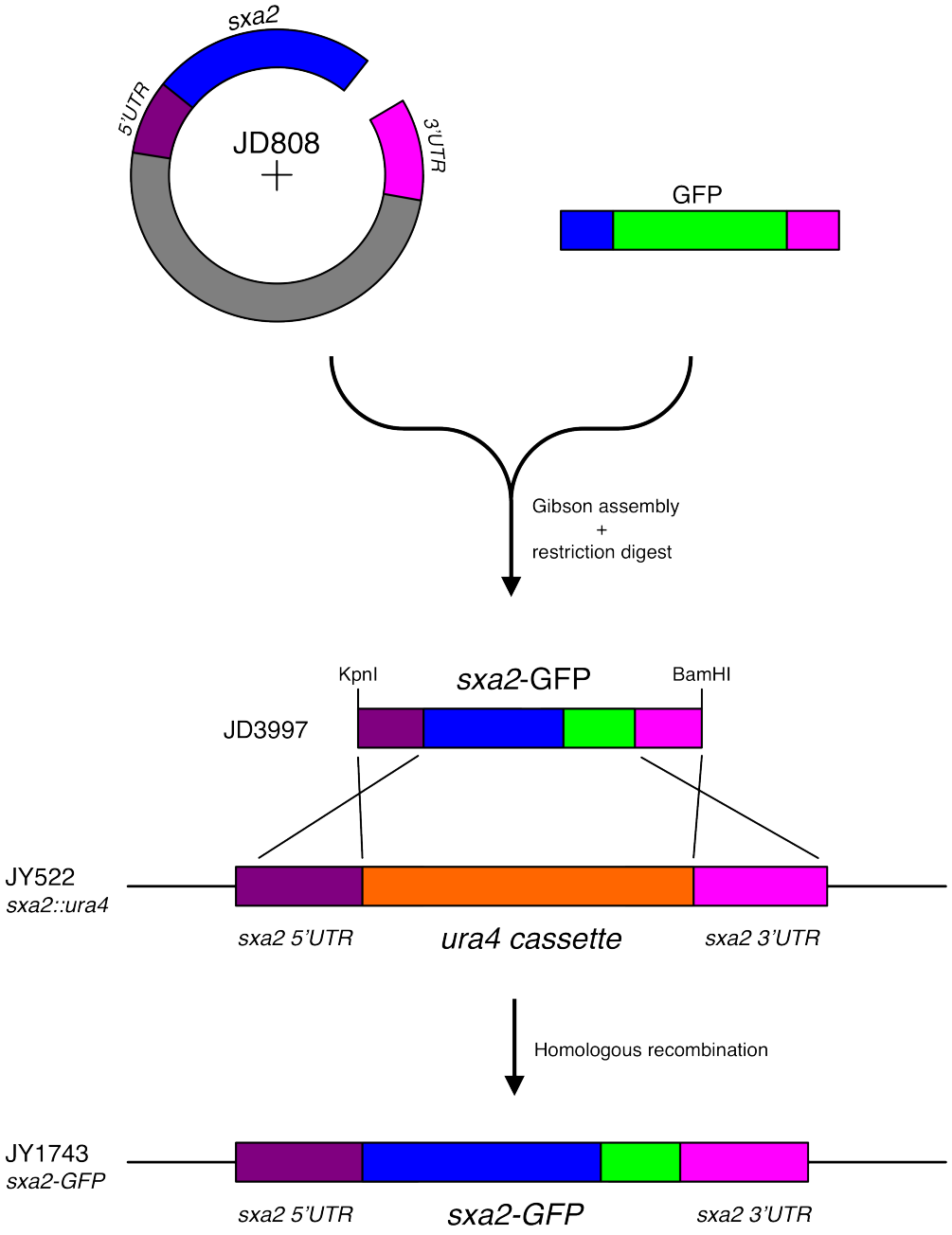


FIGURE 4.6: Creating a *sxa2-GFP* yeast strain. Plasmids JD808 containing the *sxa2* locus, and JD3387 containing a GFP reporter, were used as templates to generate DNA fragments with overlapping ends by PCR (Table 4.1). Plasmid JD3997 carrying the *sxa2-GFP* allele was created by Gibson assembly, and a DNA fragment flanked by the *sxa2* UTRs was released by restriction digest with KpnI and BamHI. Transforming strain JY522 with the *sxa2-GFP* fragment leads to replacement of the *ura4* cassette by homologous recombination.

The cloning procedure and yeast generation was facilitated by materials produced by previous projects of our research group. The yeast strain JY522 has been disrupted at the *sxa2* locus with the selectable nutritional marker *ura4* (Didmon *et al.*, 2002). This allows the creation of new *sxa2* alleles to be inserted at the endogenous locus by transforming JY522 cells with the desired DNA fragment flanked with appropriate homologous sequences, resulting in the replacement of the *ura4* cassette by homologous recombination (Grimm *et al.*, 1988). Successful replacements are then positively selected by growing cells in 5-Fluoroorotic acid (5-FOA), which will be converted into a toxic metabolite (5-fluorouracil) by cells that retain the *ura4* marker (Boeke *et al.*, 1984).

To create the *sxa2-GFP* allele flanked by *sxa2*-targetting sequences, the JD808 plasmid containing the *sxa2* open reading frame (ORF) with additional upstream and downstream flanks (Didmon *et al.*, 2002), was used to insert a GFP tag in-frame between the ORF and the 3 untranslated region (UTR), while removing the endogenous stop codon (Figure 4.6). Cloning was done by Gibson assembly, where DNA fragments with overlapping ends are joined through a single isothermal reaction that involves an exonuclease to produce 3 overhangs, a polymerase to fill gaps between annealed fragments, and a ligase to repair the nicks in the assembled DNA (Gibson *et al.*, 2009). The required DNA fragments with overlapping ends were generated by PCR using oligonucleotides designed through the NEBuilder assembly tool web application provided by the manufacturer of Gibson assembly reagents (New England BioLabs, UK) (Table 4.1).

The resulting construct containing *sxa2-GFP* (JD3997) was then digested with KpnI and BamHI restriction endonucleases to isolate the *sxa2*-targetting fragment to be used for the *ura4* replacement of JY522 (Figure 4.6). Following the successful integration of *sxa2-GFP* at the correct locus, the resulting strain (JY1743) was tested for Sxa2 activity using the pheromone tracking assay developed previously

TABLE 4.1: Oligonucleotides for creating a *sxa2-GFP* allele. Base pairs in uppercase have identical sequence to the target. Base pairs in lower case provide overlapping ends for Gibson assembly.

Name	Sequence	Template
JO3294	ttgaaaatacgttgctttatATGAGTAAAGGAGAAGAACTTTTC	JD3387
JO3295	aaattttccgatattaaacttTTATTTGTATAGTTCATCCATGC	
JO3296	AAGTTTAATATCGGAAAATTTAAAATAC	JD808
JO3297	ATAAAGCAACGTATTTTCAAG	

(section 4.1), to verify that Sxa2-GFP exhibited the same behaviour as wild type Sxa2 (JY486).

The assay revealed that Sxa2-GFP has no catalytic activity, and was indistinguishable from a *sxa2*$^{-}$ strain; however, the lack of catalytic activity does not preclude the use of Sxa2-GFP for the stated goal of estimating the scaling parameter S_{pep}. This can still be accomplished by simultaneously fitting a wild type version of the model to wild type data, and a model without the Peptidase reaction to Sxa2-GFP data.

To ensure that the data collected from Sxa2-GFP corresponded to a truly inactive peptidase without any trace of residual activity, it was decided to generate a new *sxa2-GFP* strain carrying the S200A mutation, which completely abolishes all enzymatic activity of Sxa2 (Ladds and Davey, 2000). To create the *sxa2*^{S200A}*-GFP* allele, plasmid JD3997 was used as a template for site-directed mutagenesis (section 2.1.4), to replace the serine codon AGC for the alanine codon GCT (Table 4.2). The resulting construct (JD3998) was then used to generate a *sxa2*^{S200A}*-GFP* strain as described above. All further experiments were performed using the *sxa2*^{S200A}*-GFP* strain, so for simplicity Sxa2$^{\mathrm{S200A}}$-GFP is simply referred to as Sxa2-GFP from here on.

TABLE 4.2: Oligonucleotides for creating the S200A mutation in *sxa2*. Base pairs in uppercase have identical sequence to the target. Base pairs in lower case are complementary to each other. The base pairs that introduce the mutation are underlined.

Name	Sequence	Template
JO3242	gcgaa<u>gc</u>ttacggTAGCATATGGAGCGCC	JD3997
JO3243	ccgta<u>agc</u>ttcgcCGACAAGGTAGAGCTTTTTC	

4.2.2 Time course quantification of Sxa2-GFP production in response to pheromone

To determine the concentration factor necessary to reliably detect Sxa2-GFP produced by yeast in response to pheromone, a pilot time course was performed using two different types of centrifugal ultrafiltration filters, with a corresponding ability to concentrate samples by a factor of 10X (Amicon Ultra-0.5, Millipore, UK), from 500 µL to 50 µL, or 200X (Centricon Plus-70, Millipore, UK), from 70 mL to 350 µL. Both types of filters were chosen to have a 10,000 nominal molecular weight limit (NMWL), which is the size cut-off for particles that are filtered away, with molecules above that value will retained in the concentrated solution. The NMWL was chosen to comply with the manufacturer recommendation that for maximal recovery the NMWL should be approximately two times smaller than the molecular weight of the species of interest.

The pilot time course was performed using a pheromone dose of 1 µM, and for the 10X concentrators samples were collected and processed every hour after stimulation, for a total of 6 hrs (Figure 4.7). For the 200X concentrators only two time points were collected at 2 and 4 hrs after stimulation, due to the high cost of these disposable filters. Performing the anti-GFP ELISA on the concentrated samples showed that a 10X concentration factor was enough to detect a signal, but the

measurements were too close to the limits of the assay to provide a reliable quantification, which is recommended to be in the range between 10 to 2000 ng/mL. The results from the 200X concentration factor showed that a robust signal was generated at 4 hrs following pheromone stimulation, with a minimum acceptable reading also obtained at 2 hrs. Based on these preliminary data, the 200X concentrators were chosen for use during the final time courses.

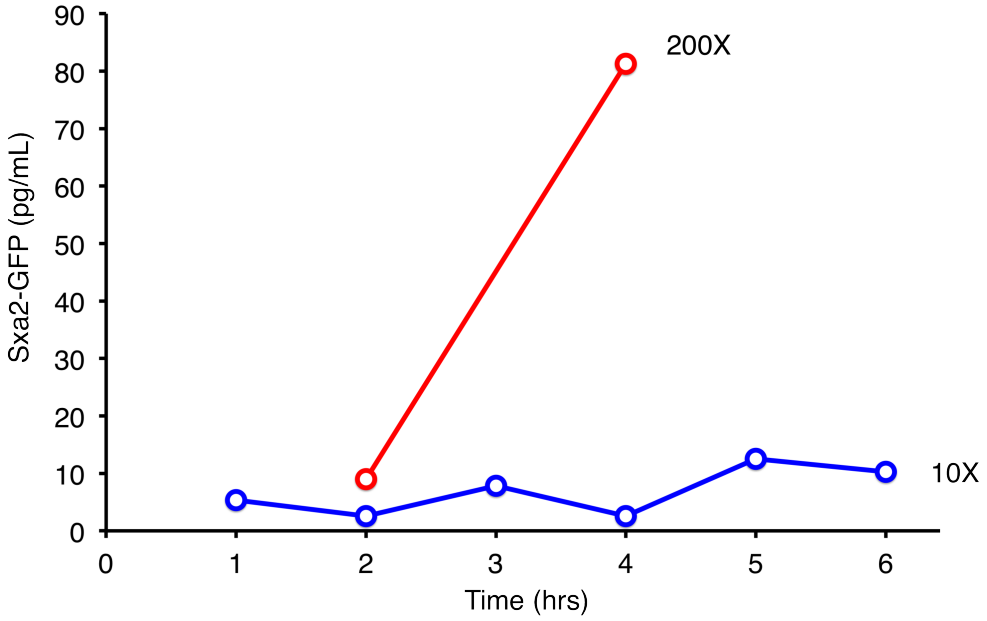


FIGURE 4.7: A 200X concentration factor is necessary to reliably quantify Sxa2-GFP. *sxa2-GFP* cells (JY1743) were cultured in DMM minimal media to mid-log phase ($\sim 5 \times 10^6$ cells/mL) and stimulated with P-factor to a final concentration of 1 µM. At the time points shown, media samples were removed and concentrated by a factor of 10X or 200X by ultrafiltration through a 10,000 molecular weight cut-off membrane. Sxa2-GFP was quantified with an anti-GFP ELISA kit (abcam ab171581).

The anti-GFP ELISA kit provided a sample of lyophilised recombinant GFP to construct a standard curve that allows the quantification of test samples to be performed on an absolute scale. However, because time course samples were processed through a filtration device, the potential loss of protein during filtration could distort the relationship between the samples and the standard curve. To circumvent this problem, the standard curve used to quantify the final time courses was prepared on the same volume as the samples, and was concentrated 200X by

ultrafiltration so that the final concentration achieved was the same as a standard curve prepared according to the manufacturers instructions. Comparing standard curves prepared by both both methods showed that protein loss due to ultrafiltration was not a significant factor affecting the assay, as both preparations produced almost identical measurements (Figure 4.8).

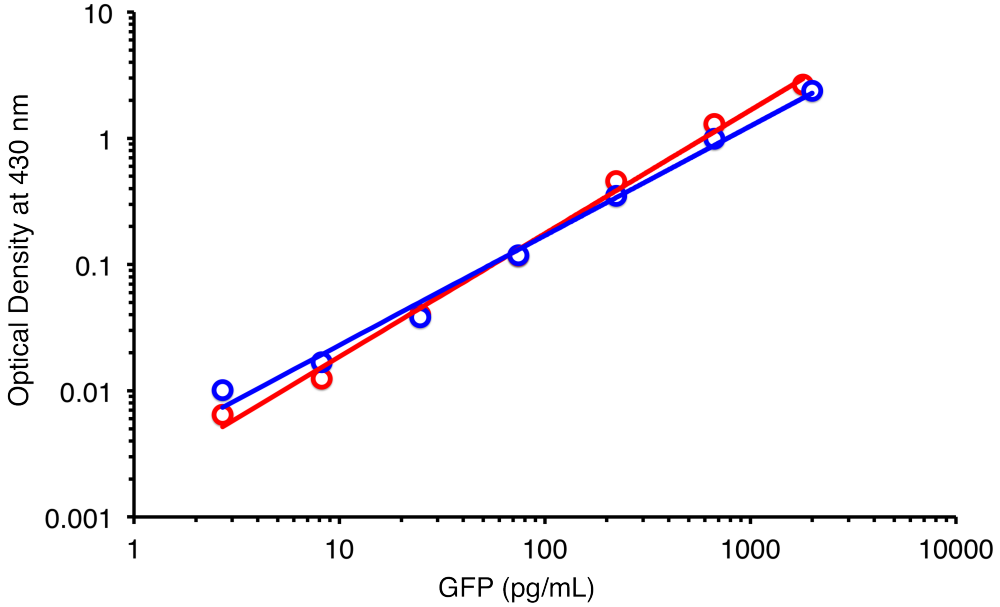


FIGURE 4.8: Protein loss due to ultrafiltration is negligible. A purified GFP standard curve was prepared in 70 mL volumes and concentrated 200X by ultrafiltration through a 10,000 molecular weight cut-off membrane down to 350 µL (Red). A similar standard curve was prepared directly into 350 µL volumes (Blue). Both standard curves were compared with an anti-GFP ELISA kit (abcam ab171581).

The final time courses were restricted to a single dose of 1 µM, and sampling was performed from 2 to 4 hrs after pheromone stimulation, at half hour intervals (Figure 4.9). This time frame was chosen based on the pilot experiment, to guarantee that good quality measurements would be obtained, and because in a wild type strain, the plateau in Sxa2 production occurs during this time (Ladds *et al.*, 1996). The accumulation of Sxa2-GFP was found to continue at every time point with no indication of reaching a steady state, which is to be expected from an experiment where the input remains constant, as *sxa2-GFP* is unable to degrade P-factor.

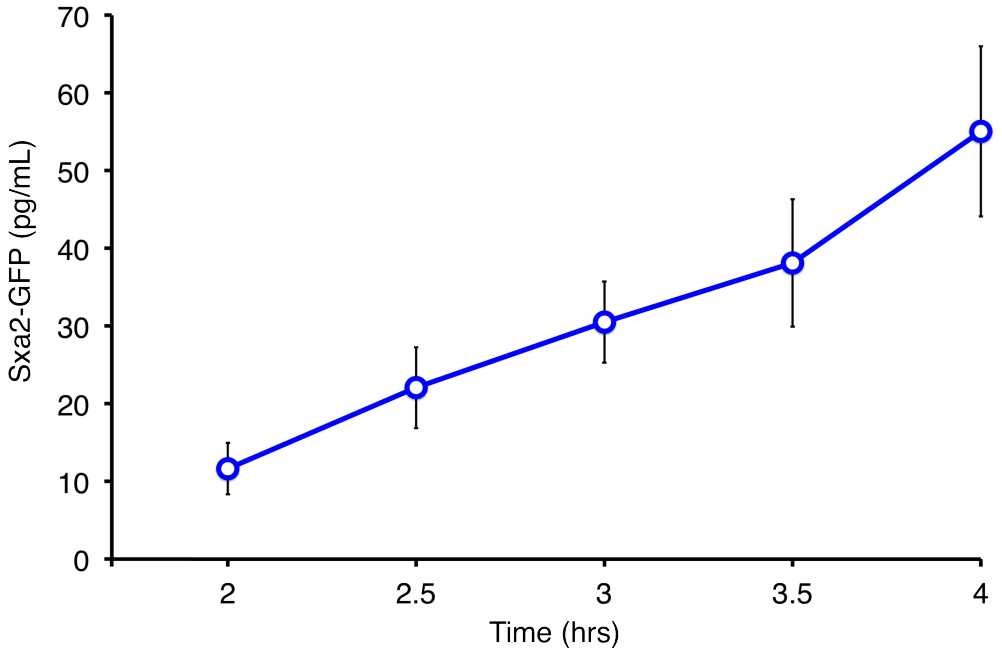


FIGURE 4.9: Absolute quantification time course of pheromone induced Sxa2-GFP. sxa2-GFP cells (JY1743) were cultured in DMM minimal media to mid-log phase ($\sim 5 \times 10^6$ cells/mL) and stimulated with P-factor to a final concentration of 1 µM. At specified time points 70 mL samples were removed and concentrated 200X by ultrafiltration through a 10,000 molecular weight cut-off membrane. Sxa2-GFP was quantified with an anti-GFP ELISA kit (abcam ab171581), using a standard curve prepared as described in Figure 4.8. Results shown are means ± SD of three biological replicates.

4.3 Quantitative gene expression analysis of *sxa2* and *ste11*

The base model A1B1C1D1E1 (3.1)-(3.5) does not consider mRNA species explicitly, however, studying the transcription dynamics of *sxa2* and *ste11* in response to pheromone was still considered an informative measurement to perform as it could be used to inform models of class A2 (Table 3.2). Among existing methods to quantify changes in transcription, the use of reverse transcription (RT) followed by real-time quantitative PCR (qPCR), termed RT-qPCR, has become the de facto standard for gene expression analysis in studies with a small number of target genes, as well as a validation tool for high-throughput technologies such as DNA microarrays (VanGuilder *et al.*, 2008).

The expression of *sxa2* and *ste11* in response to pheromone has been reported before using Northern blotting or microarrays (Sugimoto *et al.*, 1991; Imai and Yamamoto, 1992; Mata *et al.*, 2002; Xue-Franzén *et al.*, 2006; Mata and Bähler, 2006); however, the purpose of these studies was to determine differentially expressed genes, and the small number of time points analysed makes these data unsuitable for the fitting of ODE models. Here, it was sought to use RT-qPCR to generate time-resolved quantitative measurements of *sxa2* and *ste11* expression in response to pheromone, to be used for parameter estimation.

For the qPCR assays, a double-stranded DNA (dsDNA) binding dye detection chemistry was selected, due to its comparably low cost (section 2.1.9), which maximised the number of samples that could be tested. In addition, it has been reported that these dyes produce comparable results to more expensive sequence specific chemistries (Arikawa *et al.*, 2008), which gave some confidence that the quality of the data would not be compromised.

4.3.1 Reference gene selection

To make valid comparisons between RT-qPCR measurements, samples must be normalised to a reference that is assumed to be constant in all reactions (Bustin, 2000). One of the preferred choices of reference is the selection of a secondary gene target in the sample whose expression does not vary across the experimental conditions tested. The use of an internal reference gene has the benefit of being subjected to identical sources of variation as the gene of interest, such as quantity and quality of starting materials, as well as differences in RNA preparation and cDNA synthesis (Radonić *et al.*, 2004).

To select appropriate reference genes for use in pheromone stimulation experiments, the *S. pombe* genome annotation database, PomBase (Wood *et al.*, 2012), was queried for genes that showed stable expression after pheromone treatment,

as well as during environmental stress and across the cell cycle (Mata *et al.*, 2002; Chen *et al.*, 2003; Rustici *et al.*, 2004; Mata and Bähler, 2006). From the resulting list of candidates, four genes which were identified as common reference genes were selected for further validation. The selected genes were *act1* (actin), *ptr6* (transcription factor TFIID complex subunit), *rip1* (ubiquinol-cytochrome-c reductase complex subunit), and *srb4* (mediator complex subunit), which are all involved in essential cell maintenance processes, and were thus expected to maintain a stable expression under the experimental conditions tested here.

4.3.2 Primer design

Accurate quantification of DNA by qPCR is highly sensitive to the efficiency of the reaction, of which the main determinant is primer design (Quellhorst and Rulli, 2008). The basic rules for primer design are well established; however, the number of criteria that must be weighed simultaneously often precludes an optimal sequence to be found manually. To ensure the best possible primer pairs were designed for each target gene, three different freely available software tools were used to design the qPCR primers: Primer3Plus (Rozen and Skaletsky, 1999; Untergasser *et al.*, 2007), PrimerQuest (IDT, USA), and QuantPrime (Arvidsson *et al.*, 2008). Each of these programs employs a different algorithm for sequence selection, which resulted in a unique list of candidate primers produced by each one.

To further limit the number of primer pairs to be tested experimentally, the secondary structure of the amplicons produced by each primer pair was characterised using the open UNAfold software (IDT, USA; Markham and Zuker, 2008). Amplicons were discriminated based on the extent of secondary structures that were predicted to remain stable at the primer pair annealing temperature, especially in

the regions of primer binding. Only the two best primer pairs for each target gene were then synthesised for experimental validation.

To make the final selection of primer pairs, a culture of wild type cells (JY486) was grown in DMM minimal media to mid-log phase ($\sim 5 \times 10^6$ cells/mL), and a sample was processed for qPCR as described in detail in Chapter 2. Each primer pair was assayed in five replicates, and the pairs were judged for sensitivity through their cycle of quantification (C_q), and for specificity by melting curve analysis (Figure 4.10). The best performing pairs of primers for each gene were kept for further use in RT-qPCR time courses (Table 4.3).

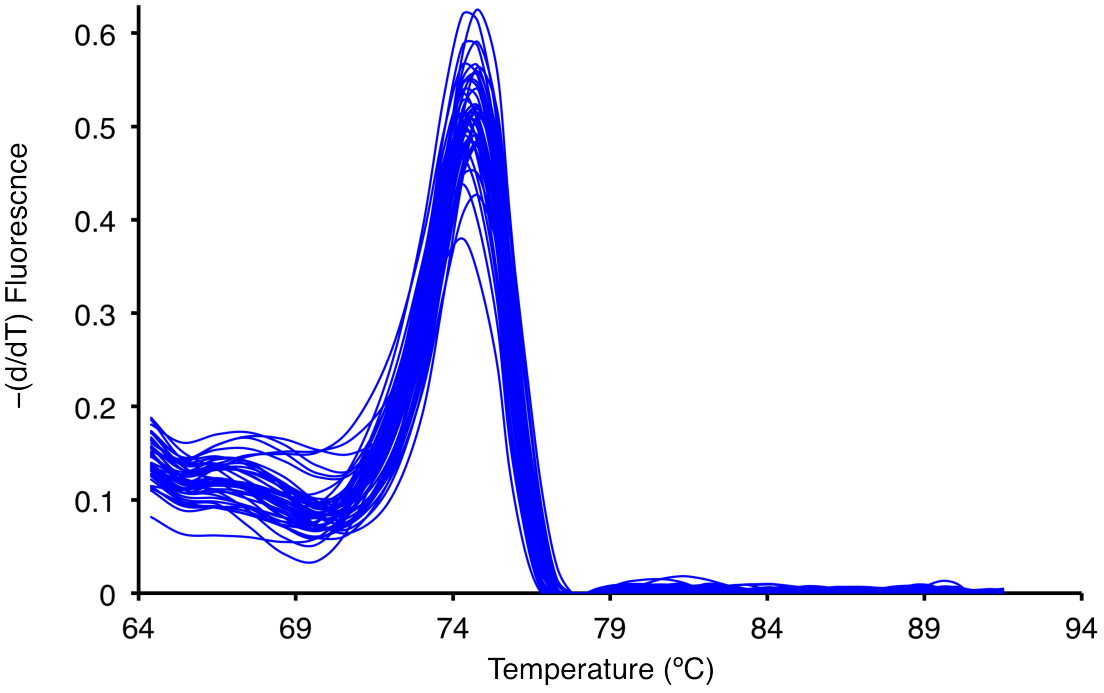


FIGURE 4.10: Post-amplification melting curve analysis. At the end of every qPCR run, each individual reaction was checked for specificity by recording fluorescence change while increasing the temperature. The rate of change in fluorescence with respect to temperature is diagnostic of the number of amplicons produced in each reaction, with a single well defined peak expected at the melting temperature of the primers.

Table 4.3: Oligonucleotide primers for RT-qPCR.

Name	Sequence	Target gene
act1F	CTCATGAGGAACTTTGGGTATCA	*act1*
act1R	GCTAGCTCTGCATTCGTCTATAA	
ptr6F	GACGACCAAAGCGTGCAAACTG	*ptr6*
ptr6R	ACGAGGTTCTGCTGTTGACTCC	
rip1F	ACTTGGCGATTCCTGCTTAT	*rip1*
rip1R	GGAGGAAATCAGCTAGGACAAA	
srb4F	TGCCCTTGATATGACCTCTTTG	*srb4*
srb4R	TGCGAACGACTAAACTGTAAGG	
sxa2F	GAATGGATGGCAGGGATTCA	*sxa2*
sxa2R	CAACCGAGTTGGACAATGTAAAG	
ste11F	CTGGCTATACATGCGTCTCTTC	*ste11*
ste11R	CGTCGCTTCTCCAGGTATTATT	

4.3.3 Reference gene stability validation

To select a reference gene from the four potential candidates described above, a 4 hr time course was performed on wild type cells (JY486) grown in DMM minimal media to mid-log phase ($\sim 5 \times 10^6$ cells/mL), and either treated with P-factor to a final concentration of 1 µM, or left untreated for comparison. Processed samples were normalised to total RNA concentration before cDNA synthesis, and equal amounts of cDNA were quantified by qPCR. All four genes showed very little discrepancy between treated and untreated samples, and only small fluctuations in their C_q values throughout the time course (Figure 4.11). To make an objective decision, the time courses were analysed using two different reference gene selection tools: BestKeeper (Pfaffl *et al.*, 2004), and Normfinder (Andersen *et al.*, 2004), with both programs agreeing that the greatest stability was displayed by *rip1*.

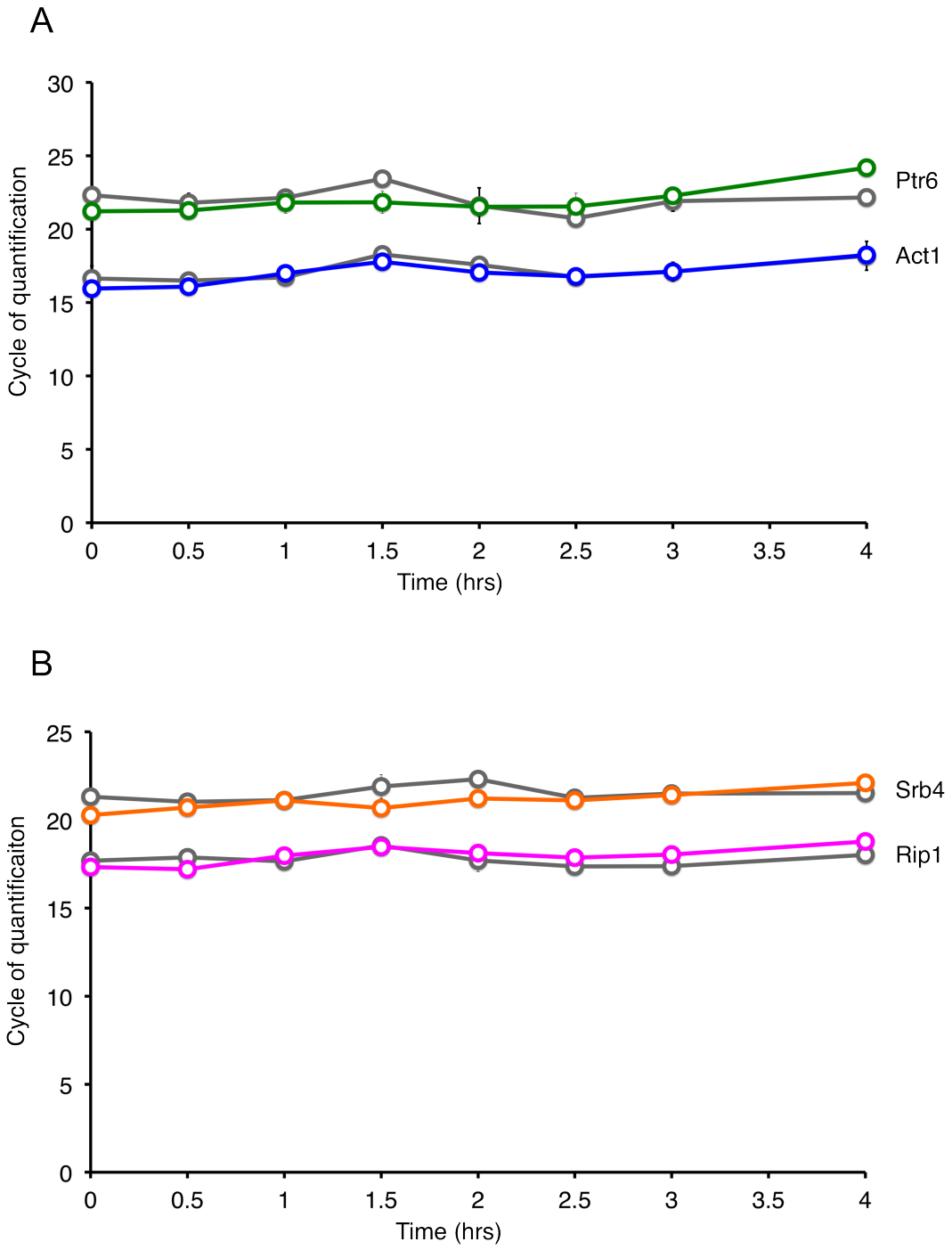


FIGURE 4.11: Validating reference gene stability under pheromone stimulation. An initial culture of wild type cells (JY486) was grown in DMM minimal media to mid-log phase ($\sim 5 \times 10^6$ cells/mL) and split into two groups. The first group was stimulated with P-factor to a final concentration of 1 µM (coloured data), and the second culture received an equivalent volume of vehicle (grey data). Samples were collected at indicated time points, and snap frozen for subsequent analysis. After RNA extraction, samples were normalised to total RNA concentration, and equal volumes of cDNA were used for qPCR.

4.3.4 Time course quantification of *sxa2* and *ste11* expression in response to pheromone

Having finalised the primer and reference gene selection, time course experiments were conducted using a final concentration of 1 µM P-factor, with samples collected every 20 min over the first 3 hrs after pheromone stimulation. This time frame was chosen because most changes in Sxa2 production occur over this period (Ladds *et al.*, 1996). To complement the inactive Sxa2-GFP protein data obtained in section 4.2 with a matching mRNA data set, time courses were also performed using a $sxa2^{S200A}$ mutant strain (JY1741), in addition to the experiments with wild type cells (JY486).

To account for possible differences in qPCR efficiency between target and reference gene reactions, the relative quantification of samples was performed using the standard curve method (Figure 4.12), where a serial dilution of pooled cDNA is analysed to determine the efficiency obtained with each primer pair, and the concentrations of test samples are interpolated from their corresponding standard curve (Pfaffl, 2001). Reaction efficiencies (E) can be estimated from the slope of the standard curves with the equation $E = -1 + 10^{-1/slope}$, and were calculated as 100.9% for *sxa2*, 100.4% for *ste11*, and 98.4% for *rip1*, where acceptable values lie between 90 and 110% (Pfaffl, 2001).

The time course data generated for *sxa2* showed distinct patterns consistent with both the pheromone tracking data (section 4.1), and the Sxa2-GFP data (section 4.2). In the wild type strain (JY486), *sxa2* levels increased steadily following P-factor exposure, reaching a maximum at around 100 min, followed by a decline for the remainder of the time course (Figure 4.13 A). The timing of this decline coincides well with the observed changes in the rate of pheromone inactivation (Figure 4.5). In contrast, for the inactive peptidase mutant (JY1741), *sxa2* levels

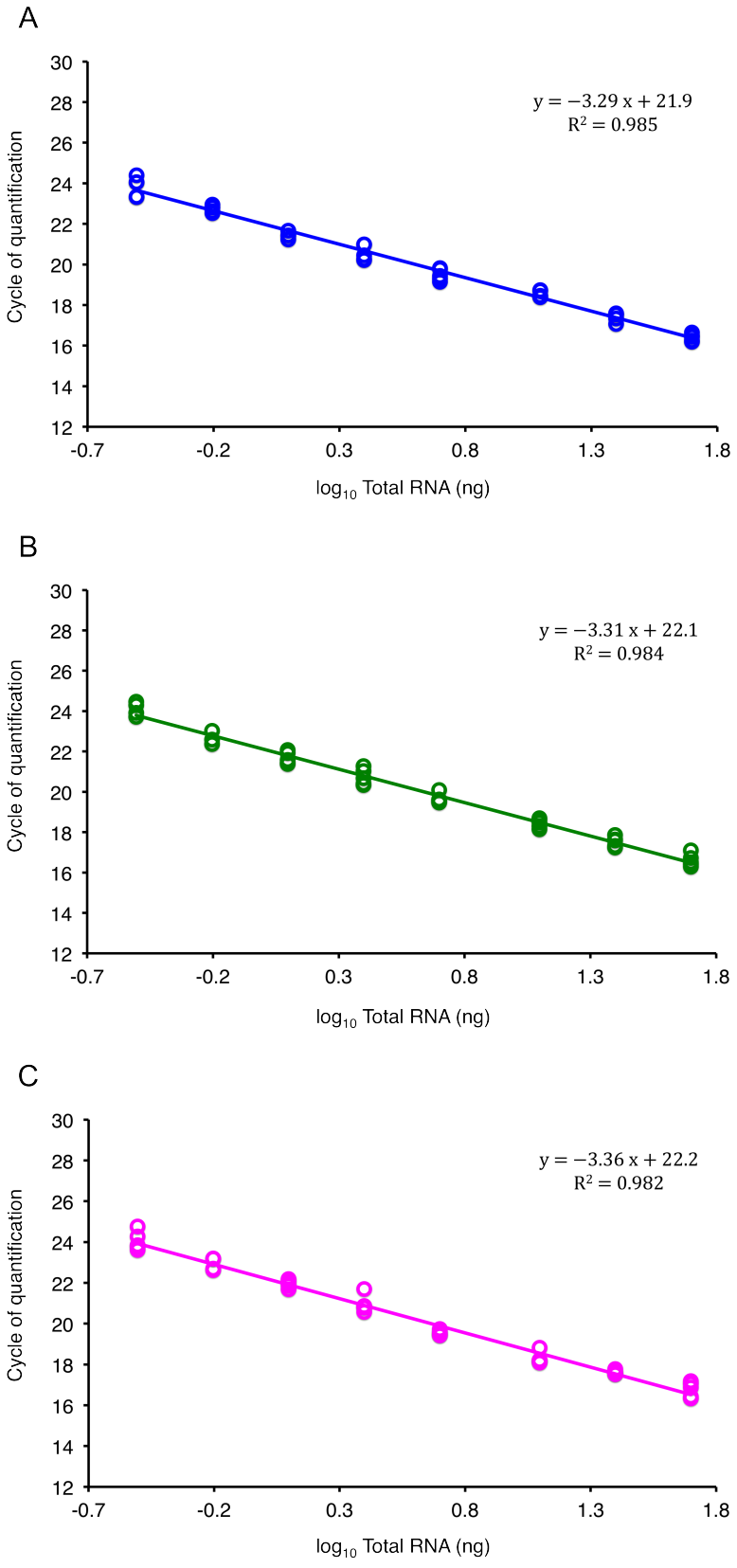


FIGURE 4.12: Standard curves for relative quantification by qPCR. To account for differences in amplification efficiency between primers, a standard curve was constructed for each biological replicate using a dilution series of pooled RNA from all time points. For each reaction, the amount of target is determined from the appropriate standard curve, followed by normalisation to the corresponding endogenous control reaction. Example standard curves are shown from one biological replicate using primers targeting *sxa2* (A), *ste11* (B), and *rip1* (C).

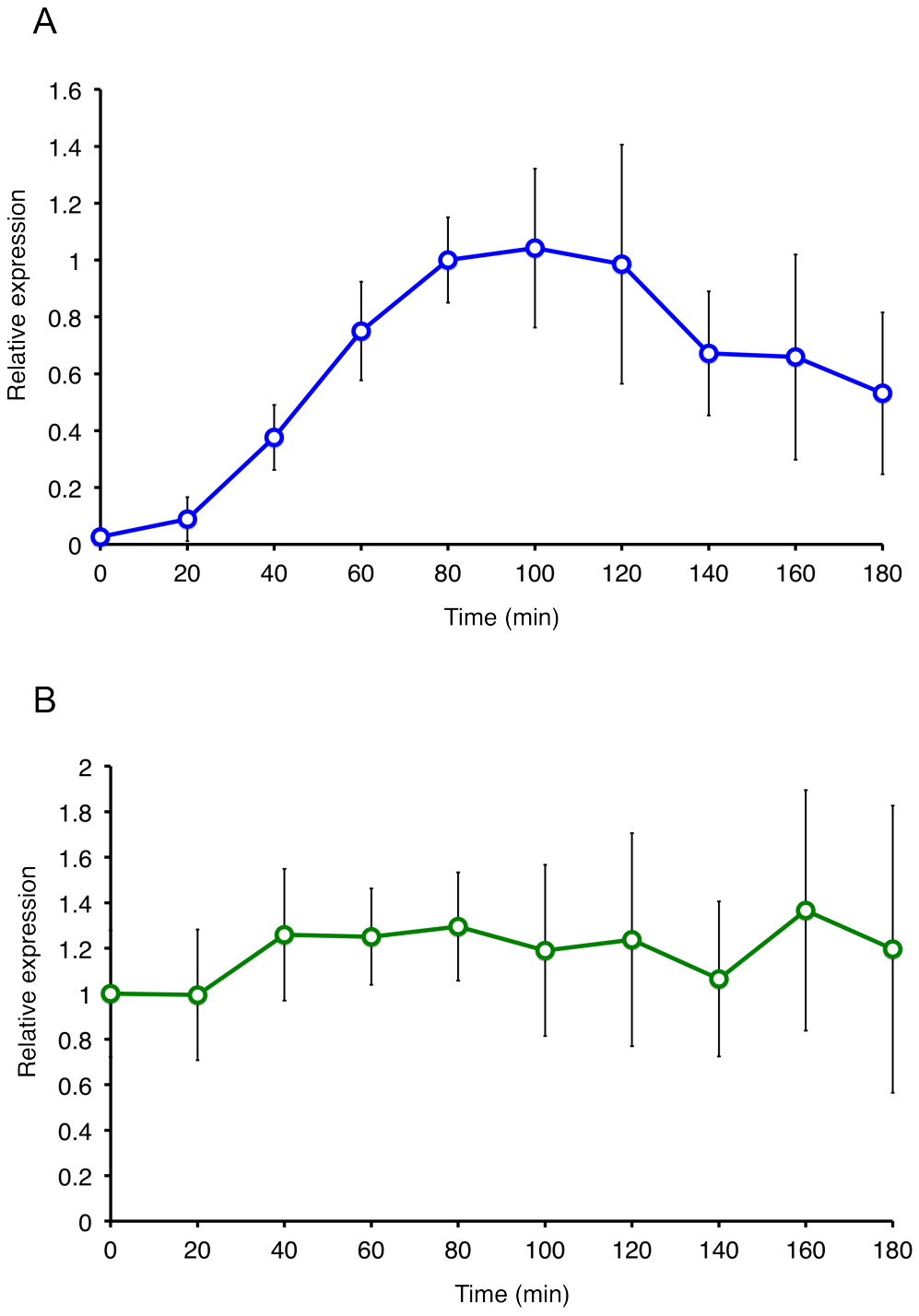


FIGURE 4.13: Relative quantification of gene expression in wild type cells responding to pheromone. Wild type cells (JY486) were cultured in DMM minimal media to mid-log phase ($\sim 5 \times 10^6$ cells/mL) and stimulated with P-factor to a final concentration of 1 µM. At specified time points, samples were collected for analysis by qPCR. (A) Time course of *sxa2*, with the 80 min time point chosen as the calibrator sample. (B) Time course of *ste11*, with the 0 min time point chosen as the calibrator sample. Results shown are means ± SD of three biological replicates.

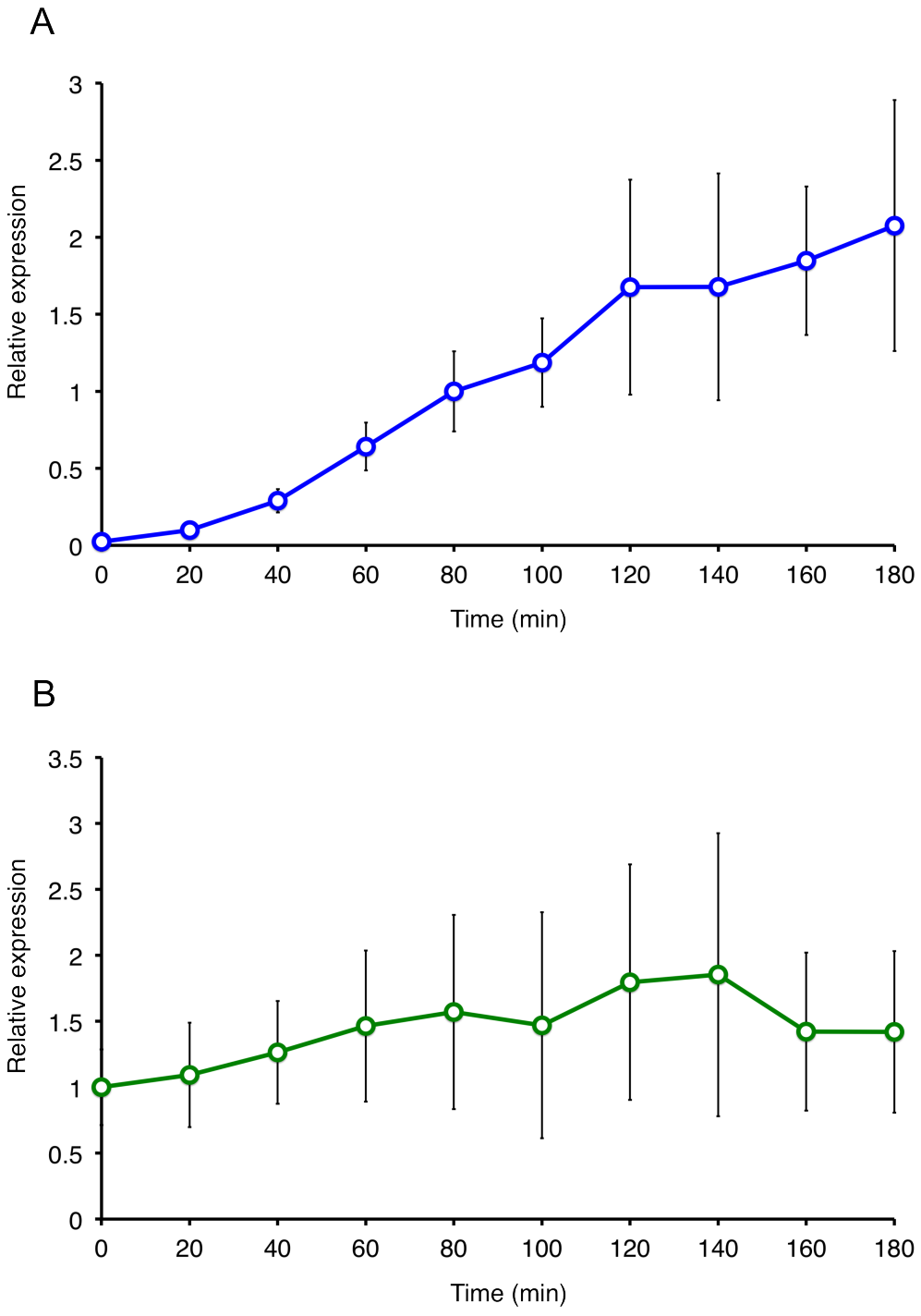


FIGURE 4.14: Relative quantification of gene expression in *sxa2*^{S200A} cells responding to pheromone. *sxa2*^{S200A} cells (JY1741) were cultured in DMM minimal media to mid-log phase ($\sim 5 \times 10^6$ cells/mL) and stimulated with P-factor to a final concentration of 1 µM. At specified time points, samples were collected for analysis by qPCR. (A) Time course of *sxa2*^{S200A}, with the 80 min time point chosen as the calibrator sample. (B) Time course of *ste11*, with the 0 min time point chosen as the calibrator sample. Results shown are means ± SD of three biological replicates.

continue to rise throughout the time course (Figure 4.14 A), in similar fashion to that observed for Sxa2-GFP (Figure 4.9).

The time course data generated for *ste11* did not display a discernible pattern due to the variability between replicate data (Figures 4.13 B and 4.14 B). The larger variability observed in *ste11* could be in part because of lower levels of expression compared to *sxa2*, as reflected in their C_q values (not shown); however, the variability in *sxa2* also appeared to increase to similar levels as *ste11* during the last hour of the experiment, even for time points with high levels of *sxa2* (Figure 4.14 A). Since the efficiency of the qPCR reactions appeared to be optimal, most of the observed variation should correspond to biological variability.

4.3.5 Absolute quantification of *sxa2* and *ste11* expression in response to pheromone

After obtaining the relative quantification results (Figures 4.13 and 4.14), the opportunity arose of retesting the samples using digital droplet PCR (ddPCR) technology. ddPCR is an emerging PCR variant for DNA quantification that is able to produce absolute quantification without the need of standard curves.

ddPCR uses microfluidics to partition a normal PCR mixture into 1 nL oil-coated droplets, which then become independent PCR reactions. Because the distribution of DNA molecules in solution during partition is random, the probability of finding a particular number of target molecules inside a droplet can be described with Poisson statistics. By scoring the droplets after PCR as either successful or failed reactions (Figure 4.15), the number of target copies in the sample can be determined from the ratio of failed/total reactions (Hindson *et al.*, 2011). The large number of droplets that are counted for each sample (∼20,000 for 20 µL reactions), results in highly precise quantification, showing a decrease in the coefficient of

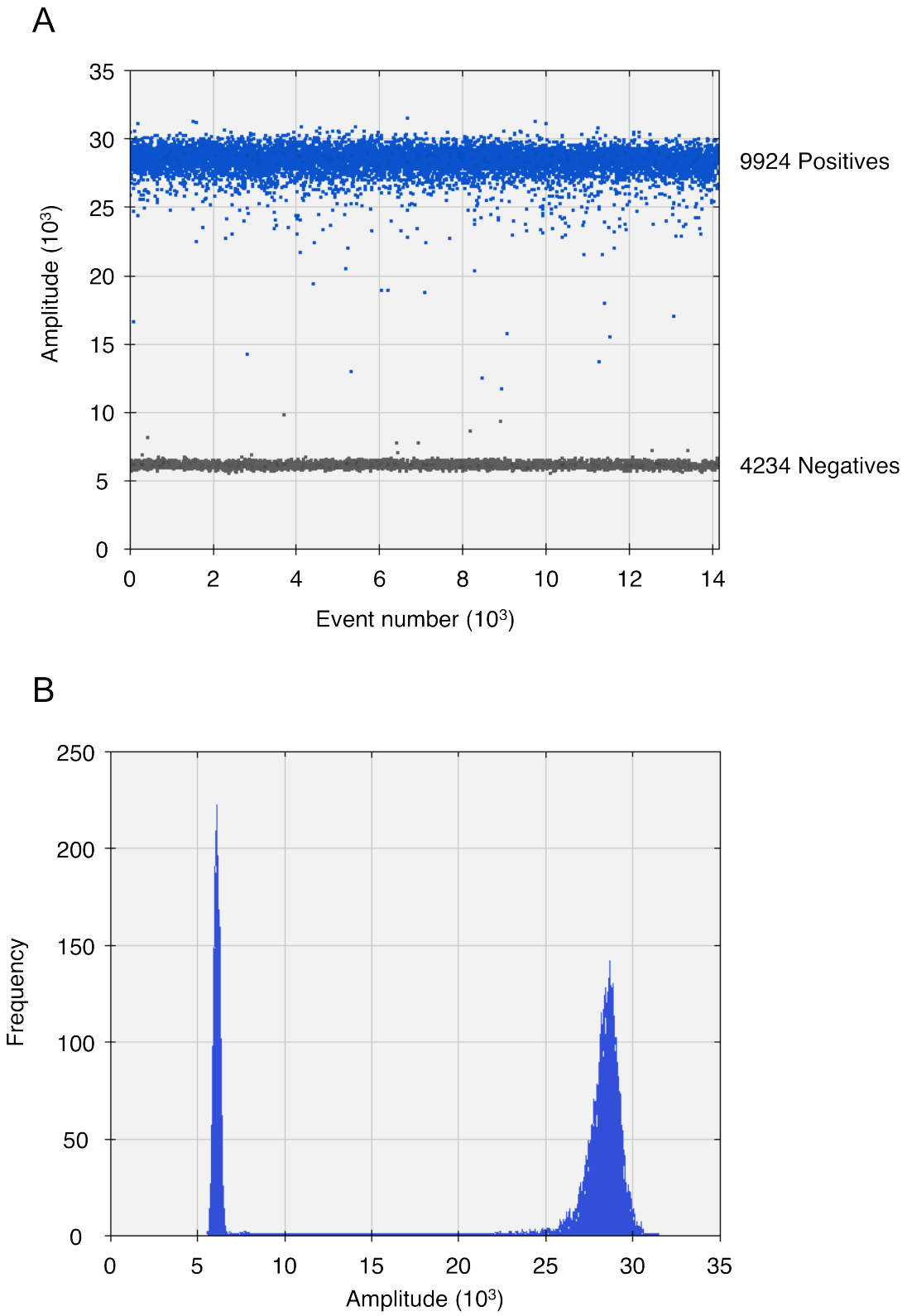


FIGURE 4.15: ddPCR provides absolute quantification without the need of a standard curve. In ddPCR, the test sample is partitioned into 1 nL droplets that may or may not contain a DNA template. Following conventional PCR with a dsDNA binding dye, each droplet is scored as either positive or negative based on fluorescence intensity. Poisson statistics can then be used to estimate the absolute initial concentration in the original sample. (A) 1-D plot of fluorescence intensity vs. droplet number. (B) Histogram of fluorescence intensity.

variation by up to 86% in comparison to qPCR, while improving inter-sample reproducibility by a factor of seven (Hindson *et al.*, 2013). In addition, since droplets are scored digitally, the error rates due to PCR efficiency bias are removed (Hindson *et al.*, 2011).

This technology has the added benefit that qPCR assays using intercalating dyes can be transferred directly to ddPCR platforms using the same primers, with only the sample dilutions requiring optimisation to ensure that cDNA concentrations are within the dynamic range of the technique (Hindson *et al.*, 2013). To determine the correct cDNA dilutions to employ for quantifications, a series of 1:1 dilutions were prepared and quantified by ddPCR (Figure 4.16). Time course samples were then appropriately diluted and quantified by ddPCR (Figures 4.17 and 4.18). To convert the resulting concentrations of copies per µL of sample to copies per cell, the known number of *rip1* mRNA molecules per cell during normal growth conditions was used for calibration (Marguerat *et al.*, 2012).

In addition to providing estimates of absolute mRNA molecule numbers, the use of ddPCR to quantify the same time course samples resulted in a marked decrease in the observed variability, which suggested that limitations intrinsic to qPCR were partly responsible for the effects shown earlier, and that the actual biological variability was much lower than anticipated. The results of *ste11* quantification with ddPCR confirmed that changes in *ste11* expression are much smaller than those observed for *sxa2*, presumably because the basal transcription rate of *ste11* in these strains is already close to the maximum rate that can be achieved from the *ste11* promoter. These results also suggested that a model variant without a positive feedback in *ste11* expression might be an adequate approximation to describe this experimental system, which was an important hypothesis to be tested in the model analysis stage.

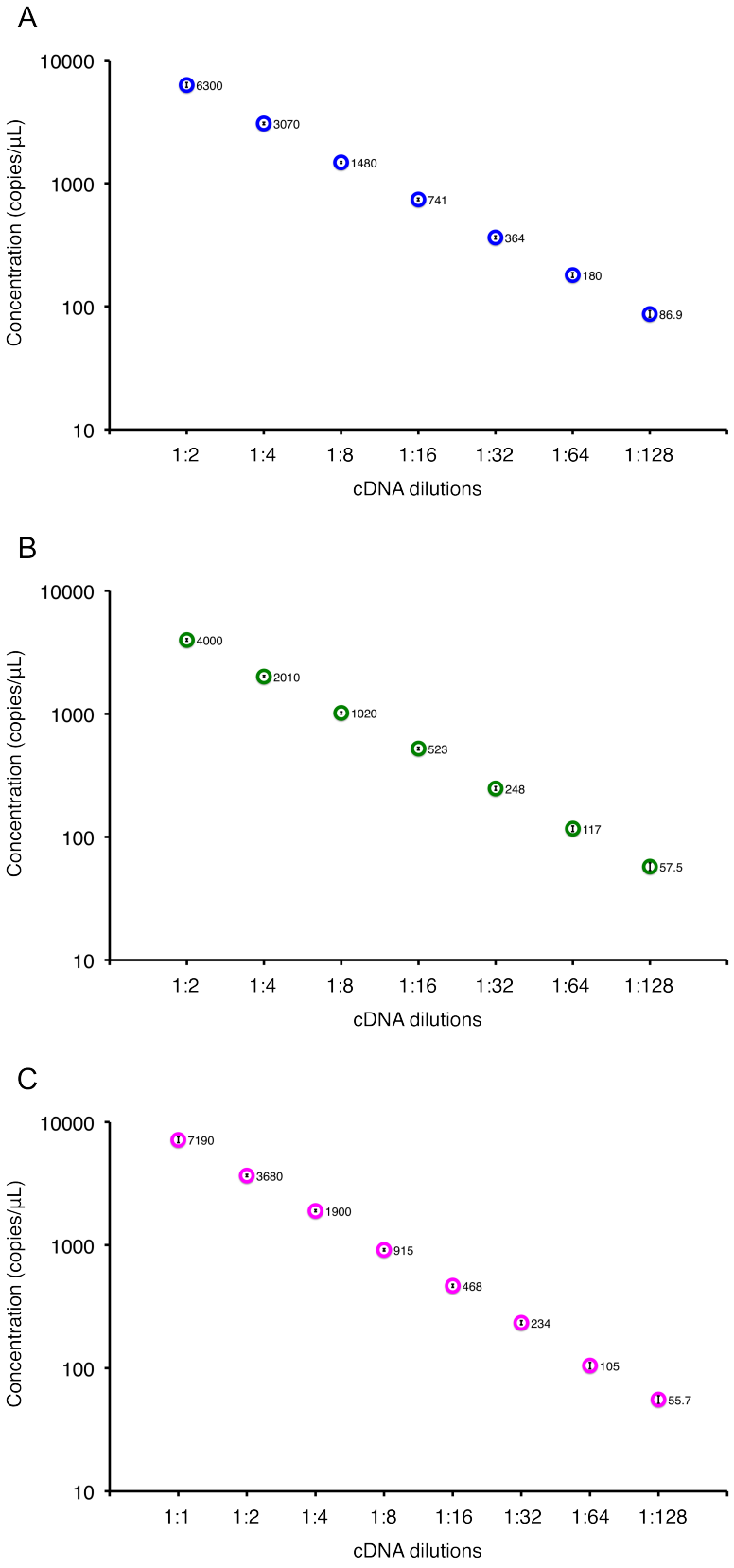


FIGURE 4.16: Establishing the correct sample dilutions for each target gene used in ddPCR. To ensure that ddPCR reactions were prepared within the dynamic range of the assay, serial dilutions of cDNA were assayed for every primer pair. (A) *sxa2*. (B) *ste11*. (C) *rip1*. Data are labelled with the estimated number of template copies per µL. Error bars are 95% confidence intervals.

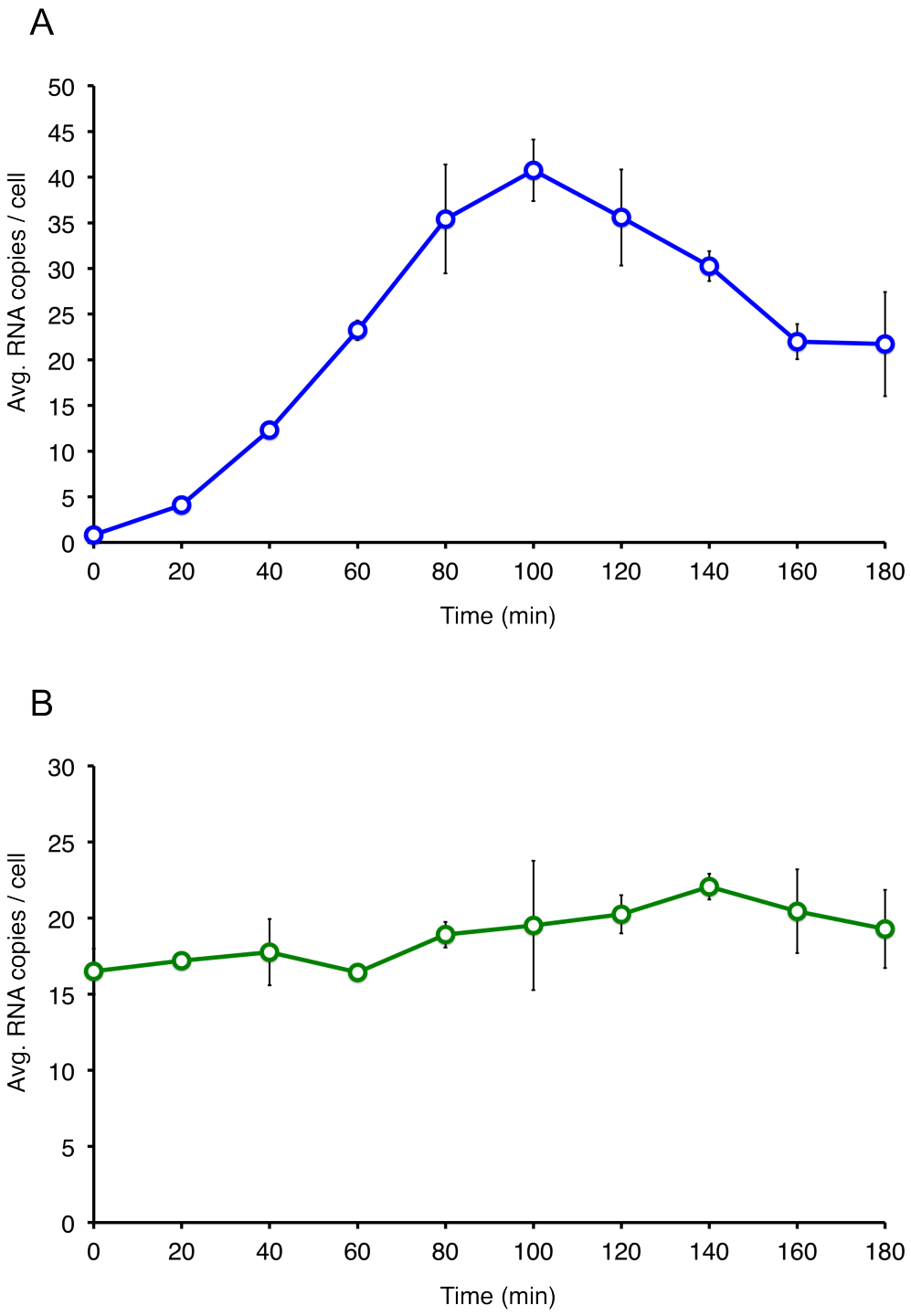


FIGURE 4.17: Absolute quantification of gene expression in wild type cells responding to pheromone. Wild type cells (JY486) were cultured in DMM minimal media to mid-log phase ($\sim 5 \times 10^6$ cells/mL) and stimulated with P-factor to a final concentration of 1 µM. At specified time points, samples were collected for analysis by ddPCR. (A) Time course of *sxa2*. (B) Time course of *ste11*. Results shown are means ± SD of three biological replicates.

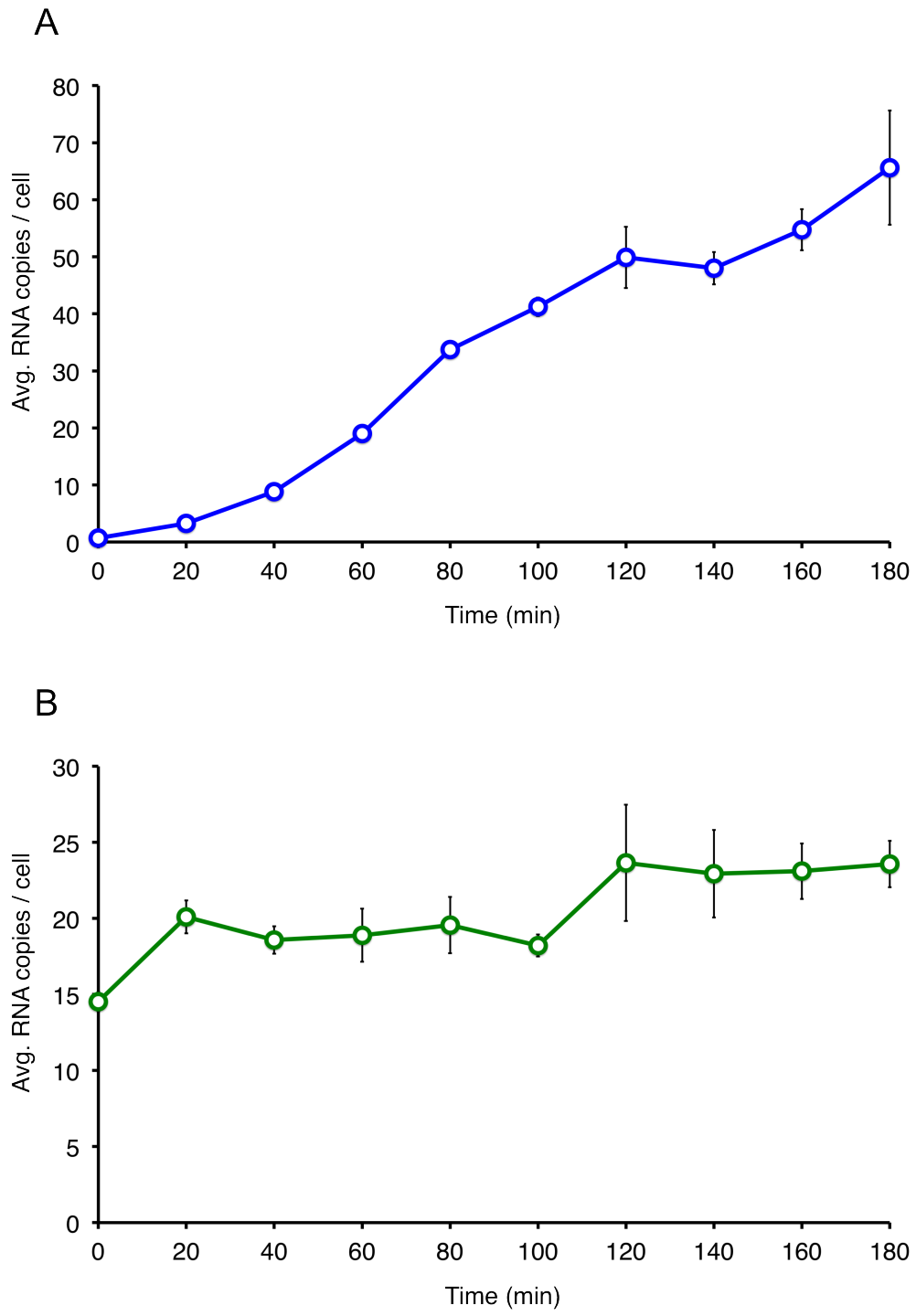


FIGURE 4.18: Absolute quantification of gene expression in *sxa2*^{S200A} cells responding to pheromone. *sxa2*^{S200A} cells (JY1741) were cultured in DMM minimal media to mid-log phase ($\sim 5 \times 10^6$ cells/mL) and stimulated with P-factor to a final concentration of 1 µM. At specified time points, samples were collected for analysis by ddPCR. (A) Time course of *sxa2*. (B) Time course of *ste11*. Results shown are means ± SD of three biological replicates.

Chapter 5

Model fitting and analysis

5.1 Fitting models to experimental data

During initial development of the model, the existing Ladds *et al.* (1996) data were used as a quick way to assess the suitability of different model variants and to gauge the complexity of the parameter estimation problem. One of the main conclusions obtained from this analysis was that using only this data set the parameter estimation problem is underdetermined, and more data were needed to reliably identify the model parameters within finite confidence intervals. Only two out of nine dynamic parameters were identifiable when fitting model A1B1C1D1E1 to the Ladds *et al.* (1996) data set: the TF degradation constant k_d, and the activation constant k_a. In addition, a scaling parameter needed to be introduced because measurements were on a relative scale.

A second problem encountered with this data set was the lack of biological replicates from which to estimate the magnitude of the experimental error. Measurement error affects the parameter estimates by weighting the squared residuals in the objective function (3.14). The standard approach to calculate the magnitude of the measurement error is to preprocess the data by estimating the mean and

variance directly from data replicates before parameter estimation. However, preprocessing is not possible for data composed of individual replicates as in the Ladds *et al.* (1996) data set. In addition, since biological experiments typically contain only a few replicates due to time and budget constraints, this situation leads to poor assessment of the true measurement error (Raue *et al.*, 2013).

Alternatively, measurement error can be estimated together with the model parameters by considering the distribution of the error as a parameterised function

$$\sigma_k(t_i, \theta) = f_{\sigma_k}\left(t_i, \boldsymbol{y}\left(t_i, \theta\right), \theta\right) \tag{5.1}$$

This approach has been demonstrated to yield statistically more accurate estimates of experimental error than the preprocessing approach, and importantly can also provide reasonable estimates for data with single replicates (Raue *et al.*, 2013).

To implement this approach in the parameter estimation procedure, measurements were considered to have additive error $y^{\dagger}_{ki} = y_{ki} + \epsilon_{ki}$, where the error term is normally distributed $\epsilon_{ki} \sim \mathcal{N}\left(0\,,\, \hat{\sigma}^2_k\left(t_i, \theta\right)\right)$. Even though non-negative data is expected to exhibit multiplicative log-normal noise, a simple log transformation of the measurements will result in additive normal noise (Kreutz *et al.*, 2007).

Choosing from the sets of time course data generated in chapter 4, models of class A1 can be fitted to the tracking of pheromone depletion (section 4.1), and to the measurements of inactive Sxa2 (section 4.2). Model A1B1C1D1E1 is shown fitted to all data sets simultaneously in Figure 5.1, including the Ladds *et al.* (1996) data.

In section 3.5, model A1B1C1D1E1 was found to be structurally globally identifiable with the observation function $y = \left[\, x_1 + x_2\,,\, x_3\,,\, x_4 \,\right]^{\mathsf{T}}$; however, it was not possible to secure any measurements for the Ste11 protein. The available measurements fitted in Figure 5.1 have the corresponding observation function

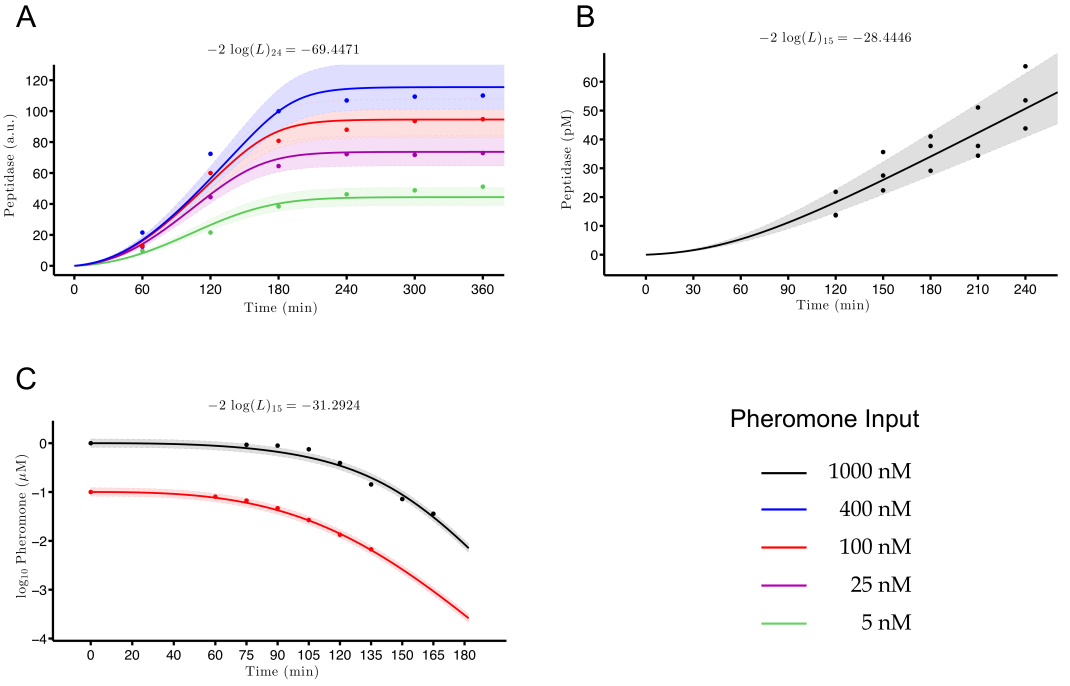


FIGURE 5.1: Fitting model A1B1C1D1E1 to experimental data. (A) Relative quantification of Sxa2 from Ladds *et al.* (1996). (B) Absolute quantification of $\text{Sxa2}^{\text{S200A}}$. (C) Absolute quantification of pheromone inactivation. Solid traces are the model trajectories fitted to experimental data (dots). Shades are the estimated experimental error of the data.

$y = [\, x_3 \,,\, x_4 \,]^{\mathsf{T}}$ which satisfy the ORC test with

$$\boldsymbol{H} = \begin{bmatrix} \mu_1 = y_1 \\ \mu_2 = y_2 \\ \mu_3 = \mathcal{L}\mu_1 \\ \mu_4 = \mathcal{L}\mu_3 \end{bmatrix} = \begin{bmatrix} x_3 \\ x_4 \\ \dfrac{v_2 x_2}{k_{m_2} + x_2} \\ -\dfrac{v_2 k_{m_2}(-k_a x_1 x_4 + k_d x_2)}{(k_{m_2} + x_2)^2} \end{bmatrix} \tag{5.2}$$

Repeating the identifiability analysis using the STAUS approach as described before (section 3.5) yields the following relationships for any pair of indistinguishable

parameter vectors

$$\begin{array}{lll} \dfrac{k_b}{k_{m_2}} = \dfrac{\bar{k}_b}{\bar{k}_{m_2}} & \dfrac{v_1}{k_{m_2}} = \dfrac{\bar{v}_1}{\bar{k}_{m_2}} & \dfrac{k_{m_1}}{k_{m_2}} = \dfrac{\bar{k}_{m_1}}{\bar{k}_{m_2}} \\ k_d = \bar{k}_d & k_a = \bar{k}_a & v_2 = \bar{v}_2 \\ k_c = \bar{k}_c & k_{m_3} = \bar{k}_{m_3} & \end{array} \tag{5.3}$$

where the parameters k_d, k_a, v_2, k_c, k_{m_3}, are identifiable, but the parameter k_{m_2} is involved in a functional relationship with parameters k_b, v_1, and k_{m_1}, rendering the model unidentifiable. This reflects the fact that k_b, v_1, and k_{m_1} are involved in the production of transcription factor TF, and our data sets lack direct information about Ste11.

The result of this analysis guarantees that it will not be possible to identify all of the model parameters uniquely by using only measurements of Sxa2 peptidase and P-factor pheromone, and at best only the parameters k_d, k_a, v_2, k_c, k_{m_3} can be identified. This was confirmed by PLE analysis, where only the parameters k_d, k_a, k_c, k_{m_3} were found to have finite confidence intervals (Table 5.1). Although the STAUS analysis shows that in principle the parameter v_2 can also be identified, the level of precision in the data did not allow v_2 to be determined in practice (i.e. numerically).

Comparing the estimated parameters in Table 5.1 with those obtained in chapter 3, shows that providing absolute quantification measurements of Sxa2 allows the scaling parameter S_{pep} to be determined, while the pheromone tracking data provide the necessary information to determine both parameters governing pheromone dynamics. It is also shown that while the new confidence interval for k_a falls entirely within the previous estimate obtained using the Ladds *et al.* (1996) data set only,

TABLE 5.1: Estimated values $\hat{\theta}_i$ of identifiable parameters with individual 95% confidence intervals $[\sigma^-, \sigma^+]$ determined by PLE for model A1B1C1D1E1 fit to experimental data in Figure 5.1. Values are given in $\log_{10}$ scale.

Parameter	$\hat{\theta}_i$	σ^-	σ^+
k_d	-1.33	-1.59	-0.89
k_a	$+1.65$	$+1.25$	$+2.36$
k_c	-2.81	-2.90	-2.69
k_{m_3}	-0.25	-0.38	-0.09
S_{pep}	$+0.49$	$+0.43$	$+0.55$

in the case of k_d there is only a partial overlap between the old and the new confidence interval, highlighting the fact that data-based identifiability methods are necessarily local.

5.1.1 Indistinguishability analysis

Although not all parameters could be identified uniquely with the set of available measurements, it may still be possible to prove that these measurements are sufficient to discriminate model A1B1C1D1E1 from alternative variants. Considering a pair of models of the form

$$\boldsymbol{\Sigma}(\boldsymbol{p}) \begin{cases} \dot{\boldsymbol{x}}(t,\boldsymbol{p}) = \boldsymbol{f}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}) \\ \boldsymbol{x}(0,\boldsymbol{p}) = \boldsymbol{x_0}(\boldsymbol{p}) \\ \boldsymbol{y}(t,\boldsymbol{p}) = \boldsymbol{h}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}) \end{cases} \tag{5.4}$$

$$\tilde{\boldsymbol{\Sigma}}(\tilde{\boldsymbol{p}}) \begin{cases} \dot{\tilde{\boldsymbol{x}}}(t,\tilde{\boldsymbol{p}}) = \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}(t,\tilde{\boldsymbol{p}}),\tilde{\boldsymbol{p}}) \\ \tilde{\boldsymbol{x}}(0,\tilde{\boldsymbol{p}}) = \tilde{\boldsymbol{x}}_{\boldsymbol{0}}(\tilde{\boldsymbol{p}}) \\ \tilde{\boldsymbol{y}}(t,\tilde{\boldsymbol{p}}) = \tilde{\boldsymbol{h}}(\tilde{\boldsymbol{x}}(t,\tilde{\boldsymbol{p}}),\tilde{\boldsymbol{p}}) \end{cases} \tag{5.5}$$

with analogous conditions to those described for model (3.24)-(3.26) in section 3.5.1. The problem of indistinguishability is concerned with establishing whether the two models (5.4) and (5.5) can be distinguished from one another on the basis of a particular set of measurements. Two models are said to be structurally indistinguishable, if for a generic $\boldsymbol{p} \in \Omega$, there exists a $\tilde{\boldsymbol{p}} \in \tilde{\Omega}$ such that $\boldsymbol{y}(t, \boldsymbol{p}) = \tilde{\boldsymbol{y}}(t, \tilde{\boldsymbol{p}})$ for all $t \in [\,0, t\,)$ (Evans *et al.*, 2004).

Since the observation functions $\boldsymbol{y}$ and $\tilde{\boldsymbol{y}}$ are unique analytic functions of time, all of their derivatives are also unique. This provides the basis for a method to assess whether two candidate models are indistinguishable (Pohjanpalo, 1978; Hattersley *et al.*, 2011).

The functions $\boldsymbol{y}$ and $\tilde{\boldsymbol{y}}$ can be represented by their Taylor series expansion around their initial condition given by

$$y(t, \boldsymbol{p}) \;=\; y(0, \boldsymbol{p}) \;+\; y^{(1)}(0, \boldsymbol{p})t \;+\; \ldots \;+\; y^{(n)}(0, \boldsymbol{p})\frac{t^n}{n!} \;+\; \ldots \tag{5.6}$$

where

$$y^{(n)}(0, \boldsymbol{p}) \;=\; \frac{d^n}{dt^n}\, y(0, \boldsymbol{p}) \tag{5.7}$$

The coefficients of these series can then be compared term by term to establish the relationships between $\boldsymbol{p}$ and $\tilde{\boldsymbol{p}}$. If at a certain point the addition of a new term results in a contradiction with respect to the relationships established from previous coefficients, then at least in principle, the two models can be distinguished from one another with the provided observations (Hattersley *et al.*, 2011; Grandjean, 2013).

Since this project was aimed at investigating the role of feedback control in the pheromone response, the most relevant variant to distinguish from the standard model was the model without positive feedback, A1B1C1D1E2. Due to the fast increasing complexity of the equations generated with this method it is not possible

to display all intermediary steps; however, it was found by comparing up to their fifth derivative, that for the two models to be indistinguishable the parameter v_1 must be equal to zero, which is generically not possible. This implies that models A1B1C1D1E1 and A1B1C1D1E2 are distinguishable with only measurements of Sxa2 peptidase and P-factor pheromone.

Fitting model A1B1C1D1E2 to all data sets simultaneously (Figure 5.2), shows that this model fails to capture several trends in the data, especially for the production of Sxa2 peptidase. While the data suggest an increase in the rate of peptidase production over time, model A1B1C1D1E2 can only fit a straight line through all data points (Figure 5.2 A and B). To compare the likelihood of each model being correct, a global AIC score was calculated for each model fitted to all three data sets, and the evidence ratio determined as described in section 3.3.

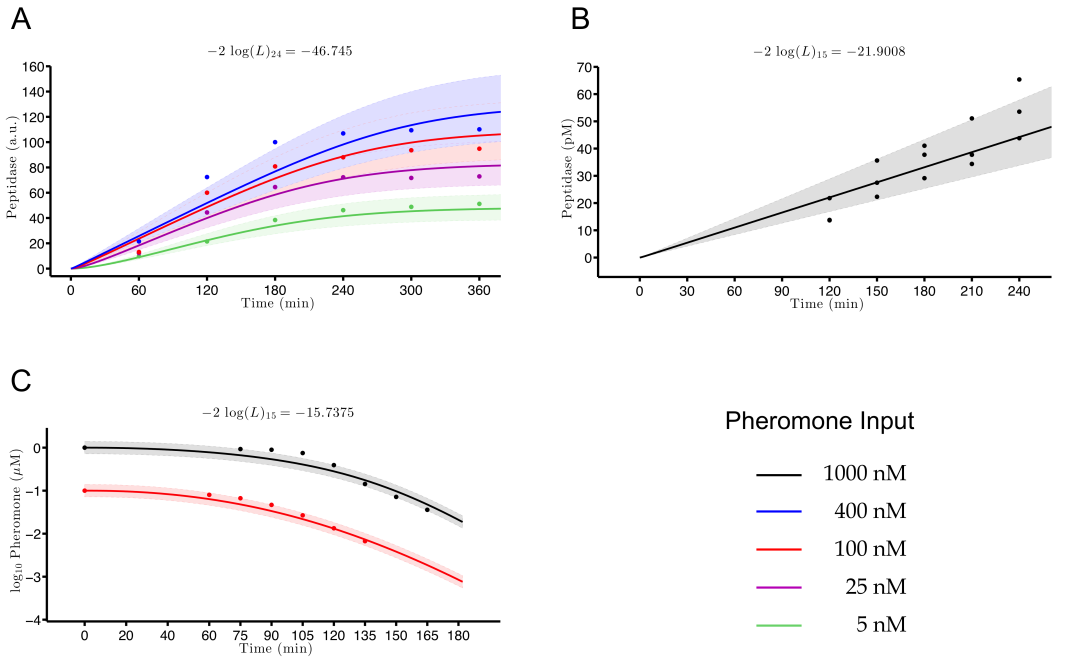


FIGURE 5.2: Fitting model A1B1C1D1E2 to experimental data. (A) Relative quantification of Sxa2 from Ladds *et al.* (1996). (B) Absolute quantification of $\text{Sxa2}^{\text{S200A}}$. (C) Absolute quantification of pheromone inactivation. Solid traces are the model trajectories fitted to experimental data (dots). Shades are the estimated experimental error of the data.

The calculated evidence ratio in favour of model A1B1C1D1E1 was 2.3×10^8, indicating that the standard model including positive feedback is overwhelmingly more likely to be correct than a model without positive feedback. However, the calculated evidence ratio may not be an accurate representation of the need to include a positive feedback in the model, since the information that distinguishes the presence or absence of positive feedback in these measurements only affects the outputs through the models' fifth derivative.

To obtain a more reliable assessment of the importance of positive feedback to describe pheromone signalling, it would be ideal to make the distinction based on a measurement that is directly affected by the feedback. In chapter 4, measurements were obtained for gene expression of both *ste11* and *sxa2* in response to pheromone stimulation (section 4.3). If there is a positive feedback that upregulates the expression of *ste11* by the activation of Ste11 transcription factors, then its effects should be most prominently observed in the data sets of *ste11* expression. Since A1 models do not consider mRNA dynamics explicitly, it is necessary to extend the analysis into A2 models.

Model A2B1C1D1E1 has the same structure as the standard model A1B1C1D1E1 and essentially describes the same system, with the addition of gene expression considered as a two-step process, with separate equations governing mRNA and protein concentrations.

The corresponding equations for model A2B1C1D1E1 are given by

$$\frac{d}{dt}\,x_1 = k_b \;+\; v_1\,\frac{x_3}{k_{m_1} + x_3} \;-\; k_{d_{m1}} \cdot x_1 \tag{5.8}$$

$$\frac{d}{dt}\,x_2 = k_{t_1} \cdot x_1 \;-\; k_{d_p} \cdot x_2 \;-\; k_a \cdot x_2 \cdot x_6 \tag{5.9}$$

$$\frac{d}{dt}\,x_3 = k_a \cdot x_2 \cdot x_6 \;-\; k_{d_p} \cdot x_3 \tag{5.10}$$

$$\frac{d}{dt}\,x_4 = v_2\,\frac{x_3}{k_{m_2} + x_3} \;-\; k_{d_{m2}} \cdot x_4 \tag{5.11}$$

$$\frac{d}{dt}\,x_5 = k_{t_2} \cdot x_4 \tag{5.12}$$

$$\frac{d}{dt}\,x_6 = -\,k_c \cdot x_5 \cdot \frac{x_6}{k_{m_3} + x_6} \tag{5.13}$$

with initial conditions at steady state given by

$$\boldsymbol{x}(0) = \boldsymbol{x_0}(\boldsymbol{p}) = \left[\begin{array}{cccccc} \dfrac{k_b}{k_{d_{m1}}} & \dfrac{k_b \cdot k_{t_1}}{k_{d_{m1}} \cdot k_{d_p}} & 0 & 0 & 0 & d \end{array}\right]^{\mathsf{T}} \tag{5.14}$$

where $\boldsymbol{x}(t) = [x_1, x_2, x_3, x_4, x_5, x_6]^{\mathsf{T}}$. The definitions of the state variables and parameters for model (5.8)-(5.14) are given in Table 5.2.

Models of class A2 can be fitted to all the data sets generated in chapter 4. The combination of all available measurements have the corresponding observation function $y = [\, x_1\,,\, x_4\,,\, x_5\,,\, x_6\,]^{\mathsf{T}}$,

TABLE 5.2: Definitions of model A2B1C1D1E1 presented in equations (5.8)-(5.14). Concentration units are in nM. Time units are in min.

State Variables	Description	Dimensions[†]
x_1	mRNA of transcription factor (Ste11)	$\mathcal{C}$
x_2	Transcription factor inactive (Ste11)	$\mathcal{C}$
x_3	Transcription factor active (Ste11)	$\mathcal{C}$
x_4	mRNA of peptidase (Sxa2)	$\mathcal{C}$
x_5	Peptidase (Sxa2)	$\mathcal{C}$
x_6	Pheromone (P-factor)	$\mathcal{C}$
Parameters		
k_b	Basal rate of TF mRNA production	$\mathcal{C}\,\mathcal{T}^{-1}$
v_1	Max. rate of TF mRNA production	$\mathcal{C}\,\mathcal{T}^{-1}$
k_{m_1}	Michaelis-Menten constant	$\mathcal{C}$
$k_{d_{m1}}$	TF mRNA degradation	$\mathcal{T}^{-1}$
k_{t_1}	TF mRNA translation	$\mathcal{T}^{-1}$
k_{d_p}	TF protein degradation	$\mathcal{T}^{-1}$
k_a	TF activation	$\mathcal{C}^{-1}\,\mathcal{T}^{-1}$
v_2	Max. rate of Peptidase mRNA production	$\mathcal{C}\,\mathcal{T}^{-1}$
k_{m_2}	Michaelis-Menten constant	$\mathcal{C}$
$k_{d_{m2}}$	Peptidase mRNA degradation	$\mathcal{T}^{-1}$
k_{t_2}	Peptidase mRNA translation	$\mathcal{T}^{-1}$
k_c	Peptidase catalytic constant	$\mathcal{T}^{-1}$
k_{m_3}	Michaelis-Menten constant	$\mathcal{C}$
d	Experimentally controlled dose	$\mathcal{C}$

[†] where $\mathcal{C}$ and $\mathcal{T}$ denote concentration and time respectively.

which satisfies the ORC with

$$\boldsymbol{H} = \begin{bmatrix} \mu_1 = y_1 \\ \mu_2 = y_2 \\ \mu_3 = y_3 \\ \mu_4 = y_4 \\ \mu_5 = \mathcal{L}\mu_2 \\ \mu_6 = \mathcal{L}\mu_5 \end{bmatrix} \tag{5.15}$$

where

$$\mu_5 = \frac{v_2 x_3 - k_{m_2} k_{d_{m2}} x_4 - k_{d_{m2}} x_3 x_4}{k_{m_2} + x_3} \tag{5.16}$$

and

$$\mu_6 = \left(k_{d_p} x_2 x_6 - k_a x_3\right) \left(\frac{v_2}{k_{m_2} + x_3} - \frac{v_2 x_3}{\left(k_{m_2} + x_3\right)^2}\right) - \left(\frac{v_2 x_3}{k_{m_2} + x_3} - k_{d_{m2}} x_4\right) k_{d_{m2}} \tag{5.17}$$

Completing the identifiability analysis using the STAUS approach yields the following relationships for any pair of indistinguishable parameter vectors

$$\begin{array}{lll} k_b = \bar{k}_b & v_1 = \bar{v}_1 & \dfrac{k_{m_1}}{k_{m_2}} = \dfrac{\bar{k}_{m_1}}{\bar{k}_{m_2}} \\ k_{d_{m1}} = \bar{k}_{d_{m1}} & \dfrac{k_{t_1}}{k_{m_2}} = \dfrac{\bar{k}_{t_1}}{\bar{k}_{m_2}} & k_{d_p} = \bar{k}_{d_p} \\ k_a = \bar{k}_a & v_2 = \bar{v}_2 & k_{d_{m2}} = \bar{k}_{d_{m2}} \\ k_{t_2} = \bar{k}_{t_2} & k_c = \bar{k}_c & k_{m_3} = \bar{k}_{m_3} \end{array} \tag{5.18}$$

where the parameters k_b, v_1, $k_{d_{m1}}$, k_{d_p}, k_a, v_2, $k_{d_{m2}}$, k_{t_2}, k_c, and k_{m_3} are uniquely

identifiable, but for parameters k_{m_1}, k_{t_1}, k_{m_2}, only the ratios $\frac{k_{m_1}}{k_{m_2}}$ and $\frac{k_{t_1}}{k_{m_2}}$ can be determined, which makes the model unidentifiable. In similar fashion to the results obtained for model A1B1C1D1E1, model A2B1C1D1E1 would be structurally globally identifiable with the addition of measurements of total Ste11 protein.

Fitting model A2B1C1D1E1 to all data sets simultaneously (Figure 5.3), it was found once more that only a subset of the theoretically identifiable parameters could be estimated within finite confidence intervals as determined by PLE analysis (Table 5.3).

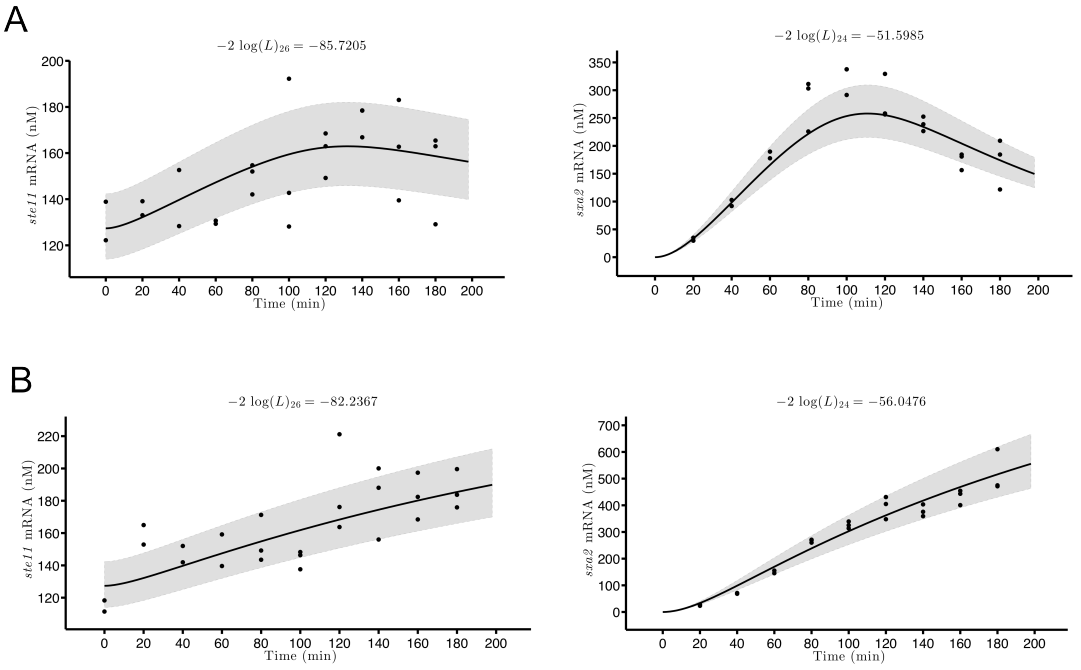


FIGURE 5.3: Fitting model A2B1C1D1E1 to experimental data. Absolute quantification of *ste11* (left) and *sxa2* (right) gene expression for the wild type strain (A), and $\text{Sxa2}^{\text{S200A}}$ mutant strain (B), stimulated with 1 µM P-factor pheromone. Solid traces are the model trajectories fitted to experimental data (dots). Shades are the estimated experimental error of the data. Only fits unique to model A2B1C1D1E1 are shown since fits to other data are essentially the same as those shown in Figure 5.1.

The estimated values and confidence intervals for parameters k_c, k_{m_3}, and S_{pep} were almost identical to those obtained with model A1B1C1D1E1. Parameters k_d and k_a on the other hand, went from being identifiable with tight confidence intervals to non-identifiable parameters. Since models A1B1C1D1E1 and A2B1C1D1E1

TABLE 5.3: Estimated values $\hat{\theta}_i$ of identifiable parameters with individual 95% confidence intervals $[\sigma^-, \sigma^+]$ determined by PLE for model A2B1C1D1E1 fit to experimental data. Values are given in $\log_{10}$ scale.

Parameter	$\hat{\theta}_i$	σ^-	σ^+
v_2	$+0.66$	$+0.59$	$+0.75$
$k_{d_{m2}}$	-1.81	-1.92	-1.71
k_{t_2}	-3.02	-3.09	-2.95
k_c	-2.83	-2.93	-2.72
k_{m_3}	-0.26	-0.39	-0.12
S_{pep}	$+0.46$	$+0.40$	$+0.51$

are representations of the same system, it is possible to use the knowledge gained from the simpler model and re-fit A2B1C1D1E1 fixing the value of parameters that were previously determined with model A1B1C1D1E1.

After fitting model A2B1C1D1E1 with constraints derived from its A1 counterpart, k_a became identifiable again, as well as parameter k_b which was previously non-identifiable for all models. The most impactful parameter to fix was k_{d_p}, since fixing this parameter yielded the same number of identified parameters as fixing all known parameters (Table 5.4).

Proceeding with the indistinguishability analysis of models A2B1C1D1E1 and A2B1C1D1E2, it was only needed to compare their first derivative to show that

$$\frac{k_b\, k_{t_1}\, d\, v_1\, k_a}{k_{m_1}\, k_{d_{m1}}\, k_{d_p}} = 0 \tag{5.19}$$

must hold for the two models to be indistinguishable, which is not possible since all parameters are generically non-zero. To provide the most stringent comparison possible, after model A2B1C1D1E2 was fitted to the data (Figure 5.4), the evidence ratio was calculated using only the AIC scores from the fit to *ste11* mRNA, because this data set is the most relevant to discriminate the effects of positive

TABLE 5.4: Estimated values $\hat{\theta}_i$ of identifiable parameters with individual 95% confidence intervals $[\sigma^-, \sigma^+]$ determined by PLE for model A2B1C1D1E1 fit to experimental data after fixing k_{d_p} (Figure 5.3). Values are given in $\log_{10}$ scale.

Parameter	$\hat{\theta}_i$	σ^-	σ^+
k_b	$+0.34$	-0.92	$+0.91$
k_a	$+1.09$	$+0.88$	$+1.25$
$k_{d_{m2}}$	-1.62	-1.76	-1.51
k_{t_2}	-3.02	-3.09	$+2.94$
k_c	-2.84	-2.93	-2.72
k_{m_3}	-0.27	-0.40	-0.13
S_{pep}	$+0.46$	$+0.39$	$+0.51$

feedback, and because this data set yielded the smallest AIC differential between the two models. Considering first the data from the wild type strain (Figure 5.4 A), the evidence ratio in favour of model A2B1C1D1E1 indicates that this model is 40 times more likely to be correct, which is equivalent to a relative likelihood of 97.5%, compared to only 2.5% for model A2B1C1D1E2. However, once the data from the mutant Sxa2^{S200A} strain are included (Figure 5.4 B), then the evidence ratio increases up to 1.2×10^6 in favour of the model including a positive feedback, which is similar to the likelihood obtained previously by comparing A1 models. Taken together, the data support a model where positive feedback to *ste11* expression is necessary to adequately explain pheromone signalling dynamics.

Further indistinguishability analyses were performed comparing the standard A1 and A2 models against other model variants; however, the computational complexity only allowed a limited number of derivative terms to be checked, and no conclusive results were found with the available measurements. It was possible to show that adding individual measurements of active and inactive TF to the set of observables would be able to distinguish between all model variants; however,

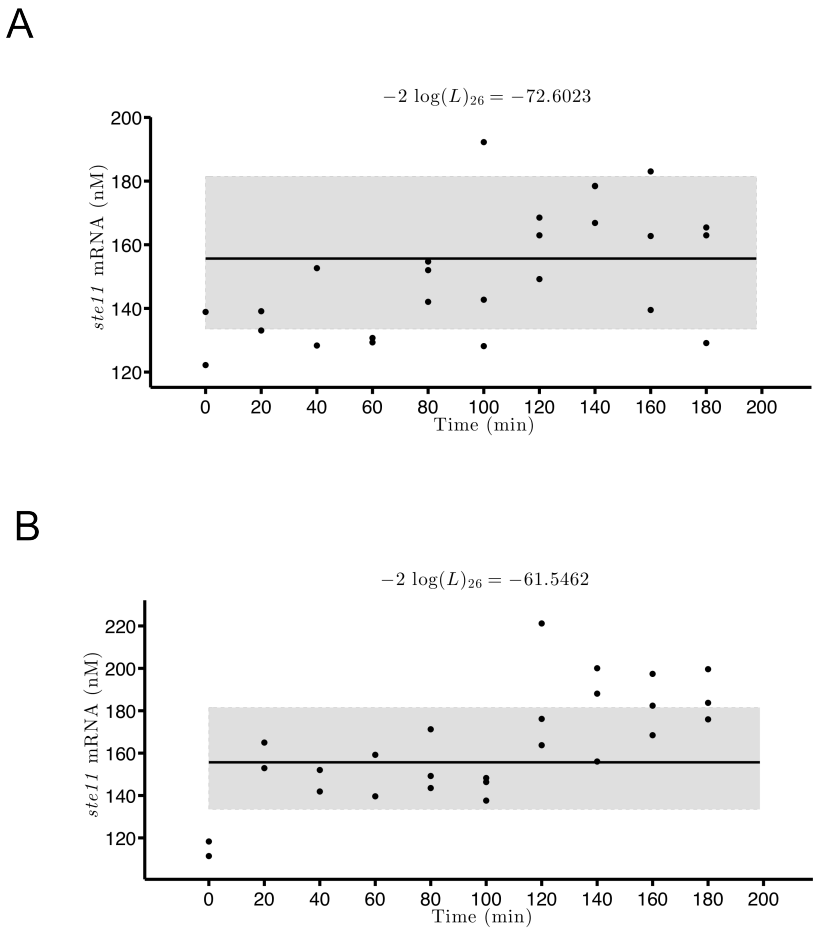


FIGURE 5.4: Fitting model A2B1C1D1E2 to experimental data. Absolute quantification of *ste11* gene expression for the wild type strain (A), and $Sxa2^{S200A}$ mutant strain (B), stimulated with 1 µM P-factor pheromone. Solid traces are the model trajectories fitted to experimental data (dots). Shades are the estimated experimental error of the data. For clarity of the argument in the main text only fits to *ste11* expression are shown.

this result is fairly trivial since it represents the case where all state variables are observable.

5.2 Experimental design to resolve parameter non-identifiabilities

Despite fitting the model A2B1C1D1E1 simultaneously to all the available data sets, several parameters remained either structurally or practically non-identifiable.

The severity of the non-identifiabilities can be assessed by inspecting the shapes of the profile likelihoods (PL) obtained through PLE analysis (Figure 5.5). The *a priori* structural identifiability analysis established that only parameters k_{m_1}, k_{t_1}, and k_{m_2} should be structurally non-identifiable and thus exhibit a flat PL, however, the PLs of $k_{d_{m1}}$ and v_2 are also completely flat, which indicates that the use of noisy data for parameter estimation renders the observables insensitive to changes in these parameters. This implies that successful identification of these parameters requires the addition of qualitatively new measurements. For the parameter v_1 only the lower bound could be determined, however, this shows that the noisy observables are not completely insensitive to changes in v_1, and that possibly a more informative experiment with the same observables could be performed.

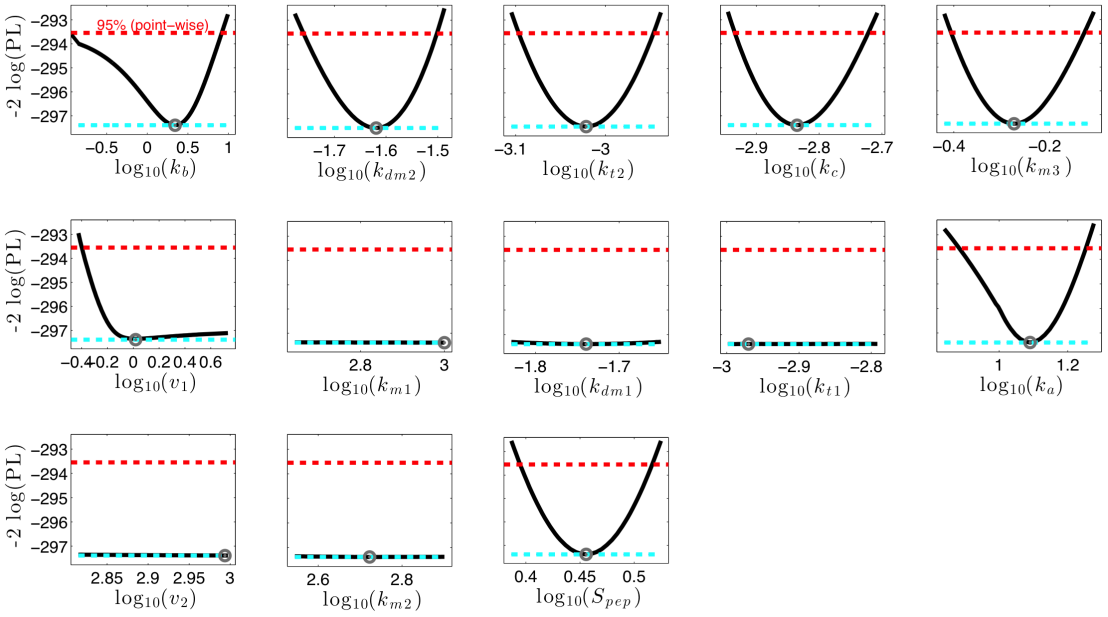


FIGURE 5.5: Parameter profile likelihoods for model A2B1C1D1E1 fitted to experimental data (section 5.1). Estimated optimum values are marked with a circle.

PLE analysis can be used to guide experimental designs either to select the most informative new species to be included, or to maximise the information that can be obtained from existing observables if qualitatively new measurements are not feasible (Steiert *et al.*, 2012). Each point along the PL of a parameter of interest

(θ_i) corresponds to a parameter vector that includes the reoptimized parameters ($\theta_{j\neq i}$). The impact of parameter uncertainty on model trajectories can be examined by simulating the model with a sample of parameter vectors along the PL, as was exemplified in Figure 3.15. The resulting trajectories reveal the extent of uncertainty in the model predictions, where a wide spread between the trajectories implies a high uncertainty and vice versa. By selecting measurements that constrain the model trajectories in the regions with most uncertainty, the resulting data will be maximally informative with respect to a reduction in parameter uncertainty (Steiert *et al.*, 2012). In addition, measurements that target regions where the uncertainty is smaller than the expected experimental error can be safely ignored, as no new information would be obtained from those data.

Since it is more difficult to generate novel measurements than it is to repeat experiments for which protocols have already been established, it was first investigated whether a resampling of existing observables could yield more informative data. The parameter v_1 is the only non-identifiable parameter that did not exhibit a completely flat PL, so the focus was placed on identifying v_1. By comparing the PL of v_1 with the changes in other parameters along PL(v_1), it can be readily seen that the parameter k_{m_1} is solely responsible for the flattening of PL(v_1) for values above the estimated optimum (Figure 5.6).

With this information in hand, the uncertainty in model trajectories was inspected for regions of shared sensitivity between v_1 and k_{m_1}. It was found that the widest spread of trajectories was in the simulation of TF mRNA, and in particular for the mutant with inactive Peptidase (Figure 5.7, see Appendix A for other observables). The trajectories begin to spread out in time points outside of the current sampling (Figure 5.7 A), suggesting that later time points might be more informative. Extending the simulations revealed a large uncertainty after doubling the time frame of the experiment (Figure 5.7 B). All other available observables displayed little

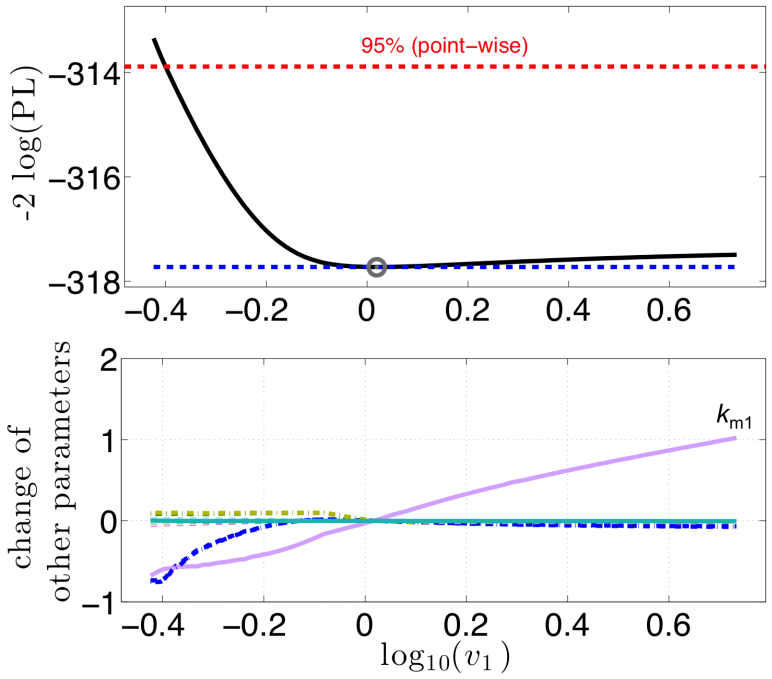


FIGURE 5.6: Non-identifiability of the parameter v_1 in model A2B1C1D1E1. (Top) Profile likelihood of v_1 from Figure 5.5. (Bottom) Changes in other parameters along the profile likelihood of v_1.

or no uncertainty with regards to v_1 and k_{m_1}, regardless of the time span allowed for simulation (Appendix A).

The impact of including measurements targeting the large uncertainty in TF mRNA at later time points can be assessed by generating synthetic data that share the same magnitude of experimental error as previously estimated for this observable during parameter estimation (section 5.1). The model can then be fitted to the extended data set (Figure 5.8 A), and the PL of v_1 is checked for improvements (Figure 5.8 B).

The parameter v_1 was found to remain practically non-identifiable despite the inclusion of additional informative data points. The PL for v_1 was only marginally improved by raising the level at which the plateau occurs (Figure 5.8 B). On the other hand the PL of k_{m_1} had a larger response to the extended data set, allowing a lower bound to be estimated, but nonetheless remaining non-identifiable (Figure 5.8 B). The fact that no additional parameters could be identified by

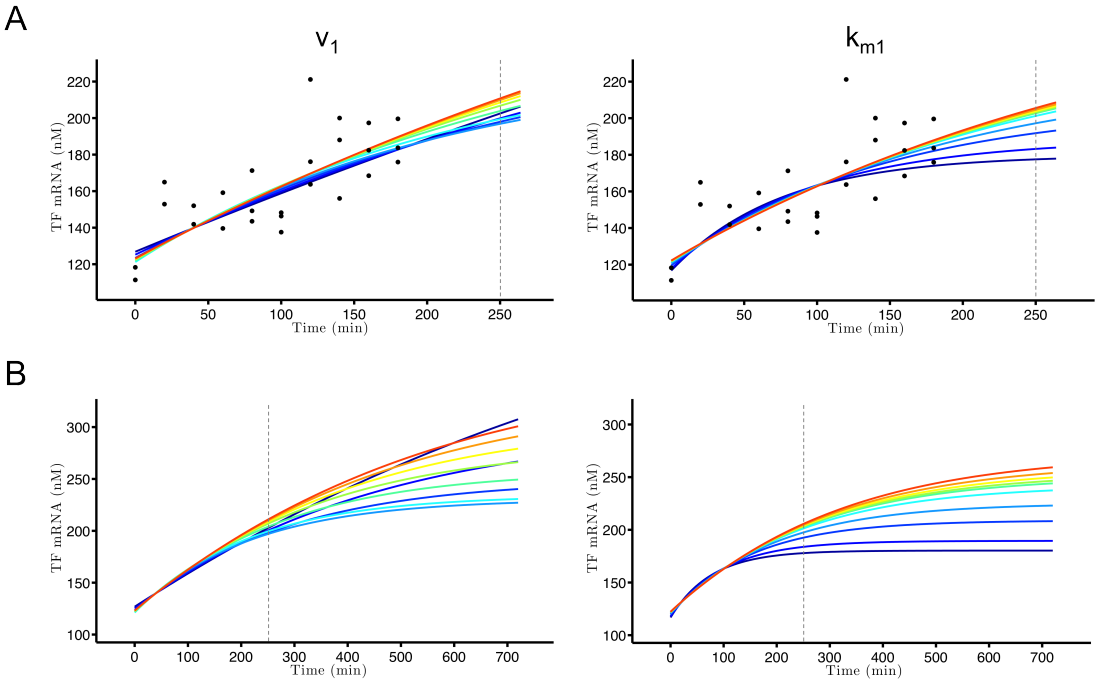


FIGURE 5.7: Uncertainty in model trajectories caused by non-identifiability of v_1 and k_{m_1}. (A) Parameter vectors along the profile likelihoods of v_1 and k_{m_1} were used to simulate the model trajectories of TF mRNA in a background of inactive Peptidase. For reference, data from this observable are also shown. (B) Extended simulation of trajectories shown in (A). For comparison, the 250 minute time point is marked in all plots.

expanding the available data sets suggests that the original time points chosen for sampling were not a factor preventing parameter identifiability, and that the maximum number possible of parameters were identified given the observables and experimental error.

To proceed with the selection of novel measurements, the uncertainty in model trajectories was inspected for all non-identifiable parameters (Appendix A). In all cases it was confirmed that no uncertainty was being propagated to the existing observables, leaving only TF inactive and TF active as potential new species to be measured; however, several redundancies limit the options even further. For the wild type condition, since TF active and TF inactive exhibit symmetrically inverted time profiles in their model trajectories, only one of the two species needs to be measured to eliminate the uncertainty in both state variables. For the inactive Peptidase mutant condition, TF active exhibits the same amount of uncertainty

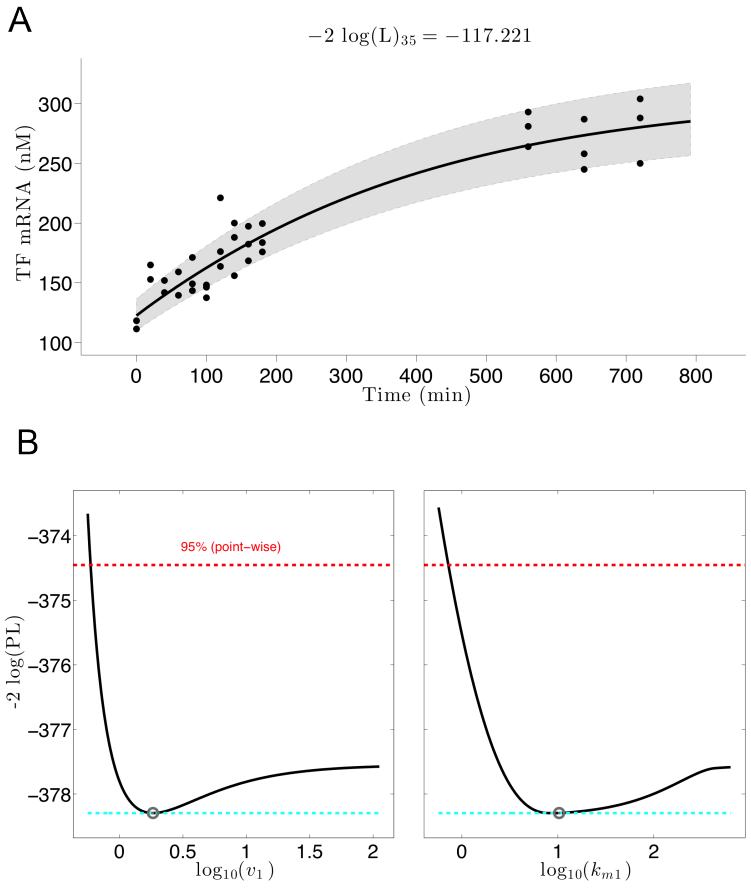


FIGURE 5.8: Effects of increased sampling from existing observables on parameter identifiability. (A) Model A2B1C1D1E1 fitted to *ste11* mRNA gene expression data from the inactive Peptidase mutant with additional synthetic data. All other data sets were also used for parameter estimation, but for clarity only the fit to the extended TF mRNA data is shown. (B) Profile likelihoods of v_1 and k_{m_1} after including additional TF mRNA data points during parameter estimation.

as the wild type case, while TF inactive provides no information because the lack of Pheromone removal in this mutant causes all of the TF pool to be in the active state at all times.

In practice, measuring active TF would be challenging since phosphorylation-specific measurements are in general harder to perform than total protein quantification. For this reason, it is worth investigating first if measurements of total TF would be a good alternative instead. This was done by defining a derived state variable consisting of the sum of both TF species, and inspecting the uncertainty

in its model trajectories to confirm that this was an informative experiment. Although not as informative as the direct measurement of the individual TF species, all non-identifiable parameters were able to affect the simulated trajectories of total TF (Appendix B). Synthetic data were then generated for total TF assuming that the experimental error would be similar to that observed when quantifying Sxa2 protein concentrations. Model fitting followed by PLE analysis revealed that including measurements of total TF is only able to resolve the non-identifiability of parameter k_{t_1} (Appendix C), leaving five undetermined parameters to be addressed.

Since measurements of total TF were not informative enough to make the model identifiable, the analysis was repeated including direct measurements of TF in either its active or inactive state, representing the case where all state variables are measured. As expected from the discussion above, measuring only one TF species had the same result as measuring both; however, having data for all species was still not sufficient to fix all the non-identifiabilities in the model. Compared to the results obtained with existing observables (Figure 5.5), including state-specific TF experiments only adds two parameters, $k_{d_{m1}}$ and k_{t_1}, to the set of identified parameters, while v_1, k_{m_1}, v_2, and k_{m_2} remain unidentifiable (Figure 5.9).

The only experimental option left to address the non-identifiabilities would be to directly measure the corresponding rate constants, or a subset of them that allows the remainder to be estimated. Inspection of the individual PL shows that the relationships preventing identification are those of v_1 with k_{m_1}, and v_2 with k_{m_2}, with no influence from any other parameters (Figure 5.10). These two relationships are independent of each other, which means that at least one rate constant from each group would need to be measured to identify all parameters.

While some parameters have a straightforward biological interpretation, such as

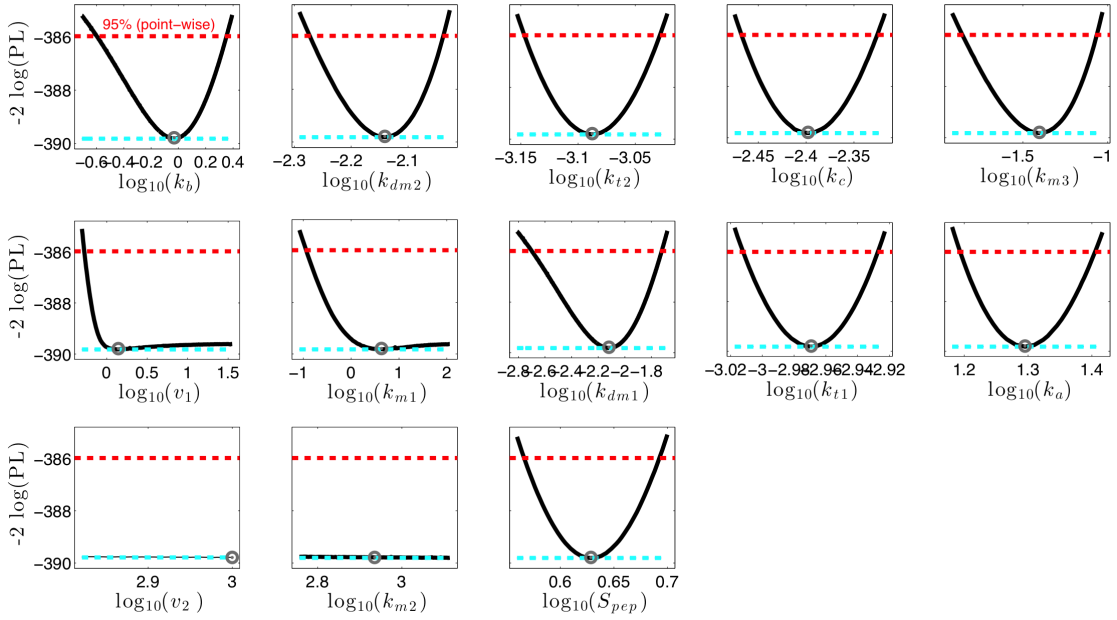


FIGURE 5.9: Parameter profile likelihoods for model A2B1C1D1E1 fitted to experimental data including synthetic data to cover all state variables. Estimated optimum values are marked with a circle.

degradation rates, the parameters v_1, k_{m_1}, v_2, and k_{m_2} are all part of the expressions used to describe gene expression. Gene expression is a multi-step process for which convenience kinetics are used, and whose parameters cannot be directly related to an individual reaction. This precludes the non-identifiability of the model to be resolved experimentally; however, a modelling alternative can be employed. Since the functional relationships are restricted to parameters within the same model reactions, i.e. TF production for v_1 and k_{m_1}, and Peptidase production for v_2 and k_{m_2}, each expression can be simplified to contain a single parameter without a significant loss in the quality of fit to the data. As it was desirable to keep the reaction rates proportional to the concentration of active TF, mass action kinetics were chosen for the replacement expressions. This alternative was not considered previously as a possible model variant in section 3.2, because saturation kinetics are the default standard used to describe gene expression in ODE models (Klipp *et al.*, 2009; Ay and Arnosti, 2011).

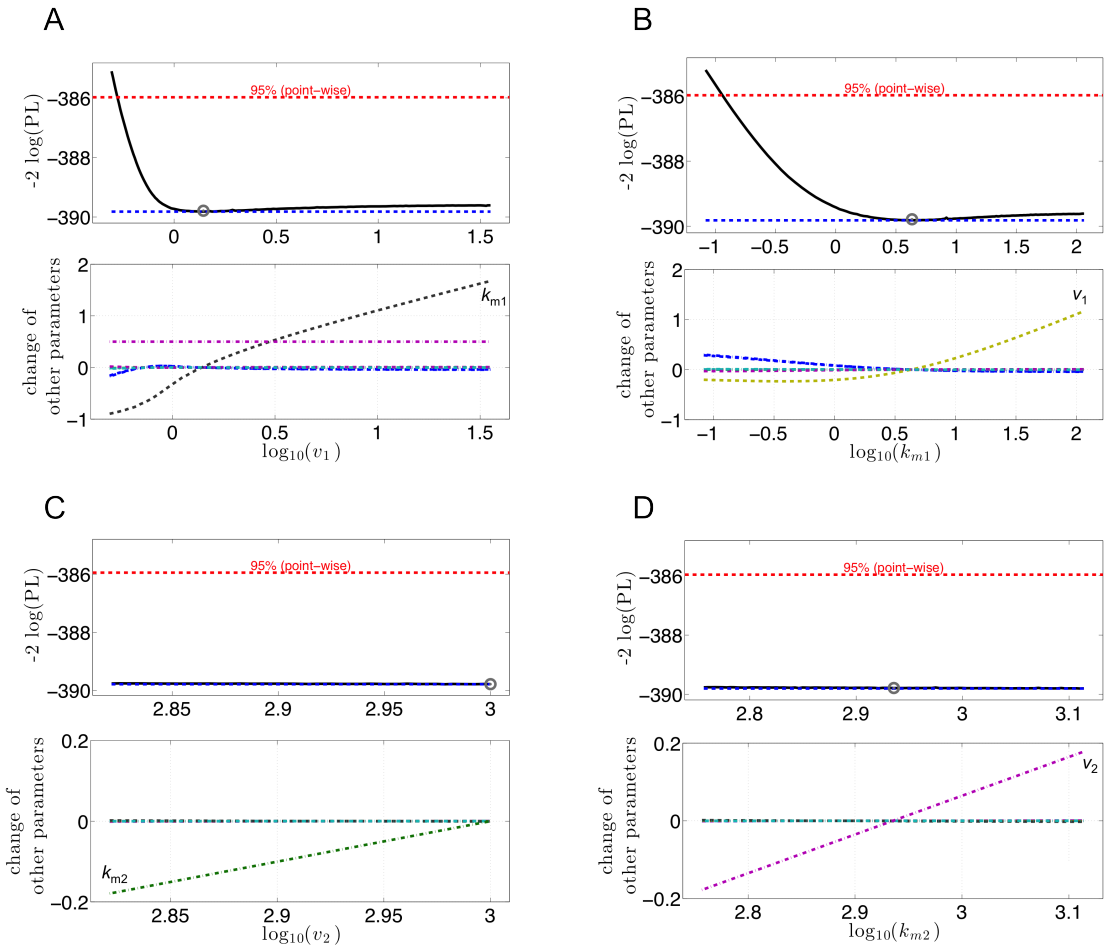


FIGURE 5.10: Non-identifiabilities of parameters in model A2B1C1D1E1 fitted to data of all state variables. For each parameter, the profile likelihood is shown together with the changes in other parameters during PLE. (A) v_1. (B) k_{m_1}. (C) v_2. (D) k_{m_2}.

To confirm that the model reduction did not significantly affect model fit, the simplified model was fitted to all data sets simultaneously (Appendix D). The resulting value of the objective function after parameter estimation was virtually the same for both models (original $2\log(L) = 297.4$; reduced $2\log(L) = 299.6$), which translated into a ΔAIC of 7.2, or a relative likelihood of $\sim$97% in favour of the reduced model. PLE analysis then showed that the same number of parameters were identifiable in both cases (Figure 5.5 and 5.11), with almost identical estimated parameter values and confidence intervals (Table 5.5). All previous results from section 5.1.1 were also found to remain valid for the model with simplified gene expression kinetics.

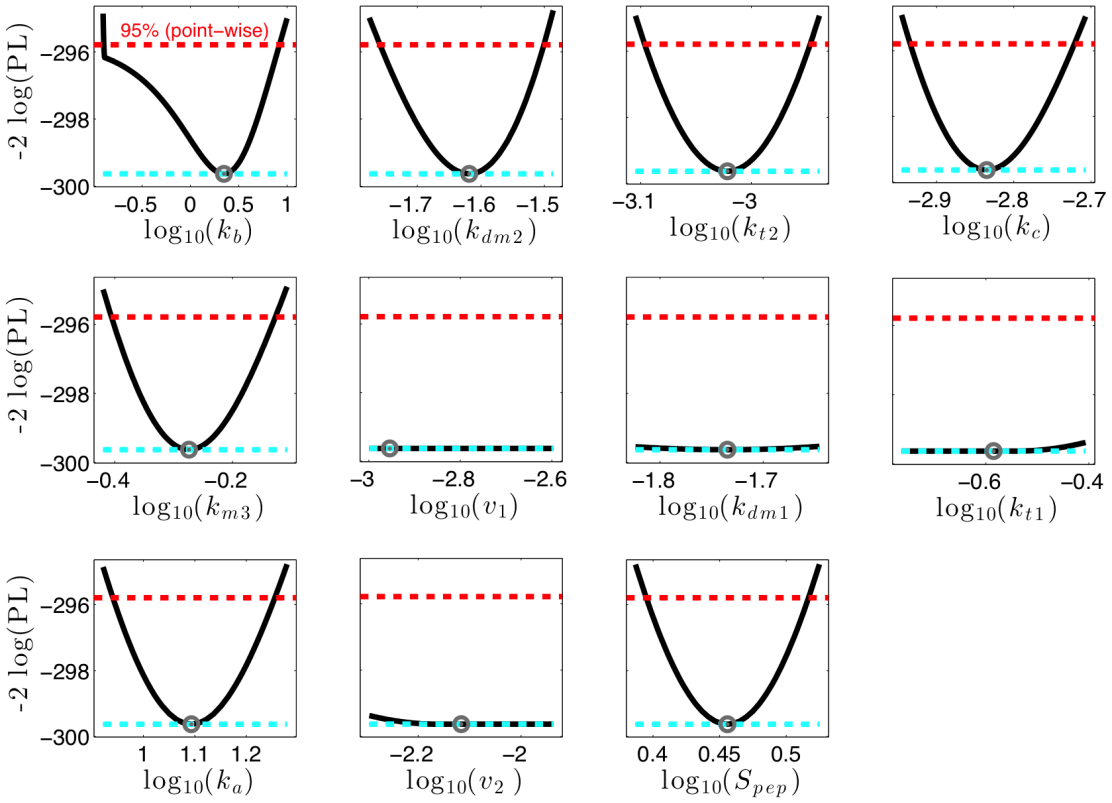


FIGURE 5.11: Parameter profile likelihoods for the reduced A2 model with simplified gene expression kinetics fitted to existing experimental data. Estimated optimum values are marked with a circle.

Having established a new standard A2 model, the methodology described in this section was applied again in search for an experimental design that could resolve all of the remaining parameter non-identifiabilities, which now consisted of only the parameters v_1, $k_{d_{m1}}$, k_{t_1}, and v_2. Extending the sampling time of TF mRNA had a greater impact than before, now managing to identify $k_{d_{m1}}$; however, no combination of increased sampling for the existing observables was capable of providing any more information. Measurements of total TF were tried next, and were found to be the only additional experiment required to render the model fully identifiable (Figure 5.12).

TABLE 5.5: Estimated parameter values and confidence intervals for the reduced model fitted to data. Estimated values $\hat{\theta}_i$ of identifiable parameters with individual 95% confidence intervals $[\sigma^-, \sigma^+]$ determined by PLE for the reduced A2 model fit to experimental data. Values are given in $\log_{10}$ scale.

Parameter	$\hat{\theta}_i$	σ^-	σ^+
k_b	$+0.35$	-0.89	$+0.92$
k_a	$+1.09$	$+0.95$	$+1.25$
$k_{d_{m2}}$	-1.62	-1.75	-1.51
k_{t_2}	-3.02	-3.09	$+2.95$
k_c	-2.83	-2.93	-2.72
k_{m_3}	-0.27	-0.41	-0.13
S_{pep}	$+0.45$	$+0.40$	$+0.51$

For conciseness, the reduced A2 model will be referred to as the A2R model from here on. For completeness, the full equations for model A2R are given by

$$\frac{d}{dt}x_1 = k_b + v_1 \cdot x_3 - k_{d_{m1}} \cdot x_1 \tag{5.20}$$

$$\frac{d}{dt}x_2 = k_{t_1} \cdot x_1 - k_{d_p} \cdot x_2 - k_a \cdot x_2 \cdot x_6 \tag{5.21}$$

$$\frac{d}{dt}x_3 = k_a \cdot x_2 \cdot x_6 - k_{d_p} \cdot x_3 \tag{5.22}$$

$$\frac{d}{dt}x_4 = v_2 \cdot x_3 - k_{d_{m2}} \cdot x_4 \tag{5.23}$$

$$\frac{d}{dt}x_5 = k_{t_2} \cdot x_4 \tag{5.24}$$

$$\frac{d}{dt}x_6 = -k_c \cdot x_5 \cdot \frac{x_6}{k_{m_3} + x_6} \tag{5.25}$$

with initial conditions at steady state given by

$$\boldsymbol{x}(0) = \boldsymbol{x_0}(\boldsymbol{p}) = \left[\begin{array}{ccccccc} \dfrac{k_b}{k_{d_{m1}}} & \dfrac{k_b \cdot k_{t_1}}{k_{d_{m1}} \cdot k_{d_p}} & 0 & 0 & 0 & d \end{array}\right]^{\mathsf{T}} \tag{5.26}$$

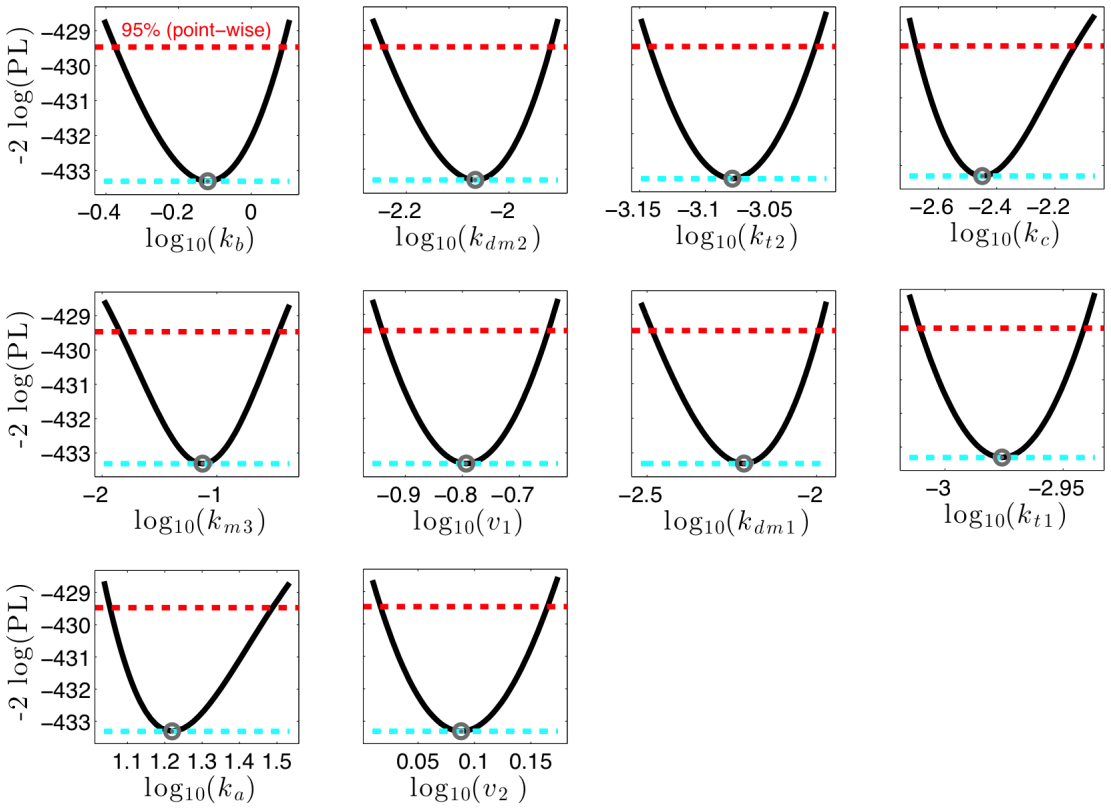


FIGURE 5.12: Parameter profile likelihoods for the reduced A2 model fitted to experimental data supplemented with synthetic measurements of total TF. Estimated optimum values are marked with a circle.

where $\boldsymbol{x}(t) = [x_1, x_2, x_3, x_4, x_5, x_6]^{\mathsf{T}}$. The definitions of the state variables and parameters for model (5.20)-(5.26) are identical to those given in Table 5.2, except for the dimensions of the parameters v_1 and v_2, which are now of inverse time.

5.3 Model analysis

After finalising the structure of the model and estimating the model parameters from available time course data, the model can then be analysed to investigate the key mechanisms that determine model behaviour, and to elucidate the roles of feedback control in this system. Although a few parameters remained unidentifiable (section 5.1), the model uncertainty derived from the non-identifiabilities

is constrained to the unobserved TF active and TF inactive species, causing a scale invariance in their concentration, and preventing quantitative predictions to be made for these species. However, the unidentifiable parameters do not affect the time scale, the shape of the trajectories, or the effects of different starting pheromone doses (Appendix A).

To provide a baseline for subsequent comparisons, the model was simulated with four different pheromone input concentrations that span the typical range used in *S. pombe* experiments (Figure 5.13), from the minimum dose that is able to provoke a measurable response, to a dose that produces a maximal response (Didmon *et al.*, 2002).

The profiles of both TF mRNA and Peptidase mRNA exhibit the same pattern, with the exception of the amplitude in their trajectories as they are plotted on different concentration scales (Figure 5.13, A and D). It is likely that any gene that depends exclusively on Ste11 for its transcription will display a similar profile, which is simply a smoothed version of the TF active profile, with an amplitude that will vary according to its own promoter efficiency. Pheromone dose affects the speed and extent of TF activation, and for a saturating pheromone dose, all of the TF pool is activated within five minutes of pheromone exposure (Figure 5.13 C). Extracellular pheromone is depleted much faster than any other intracellular species (Figure 5.13 F), which suggests that the amount of time that is available for the cell to commit to the sexual developmental programme is not only determined by the speed at which the pheromone input can be inactivated, but also by the degradation rates of regulatory mRNAs and proteins.

To determine the effects of feedback regulation on the model, the reactions responsible for the feedback can be selectively modified to alter the strength of the regulation. The resulting simulations can then be compared with the wild type version of the model to highlight their differences and to assess the impact on

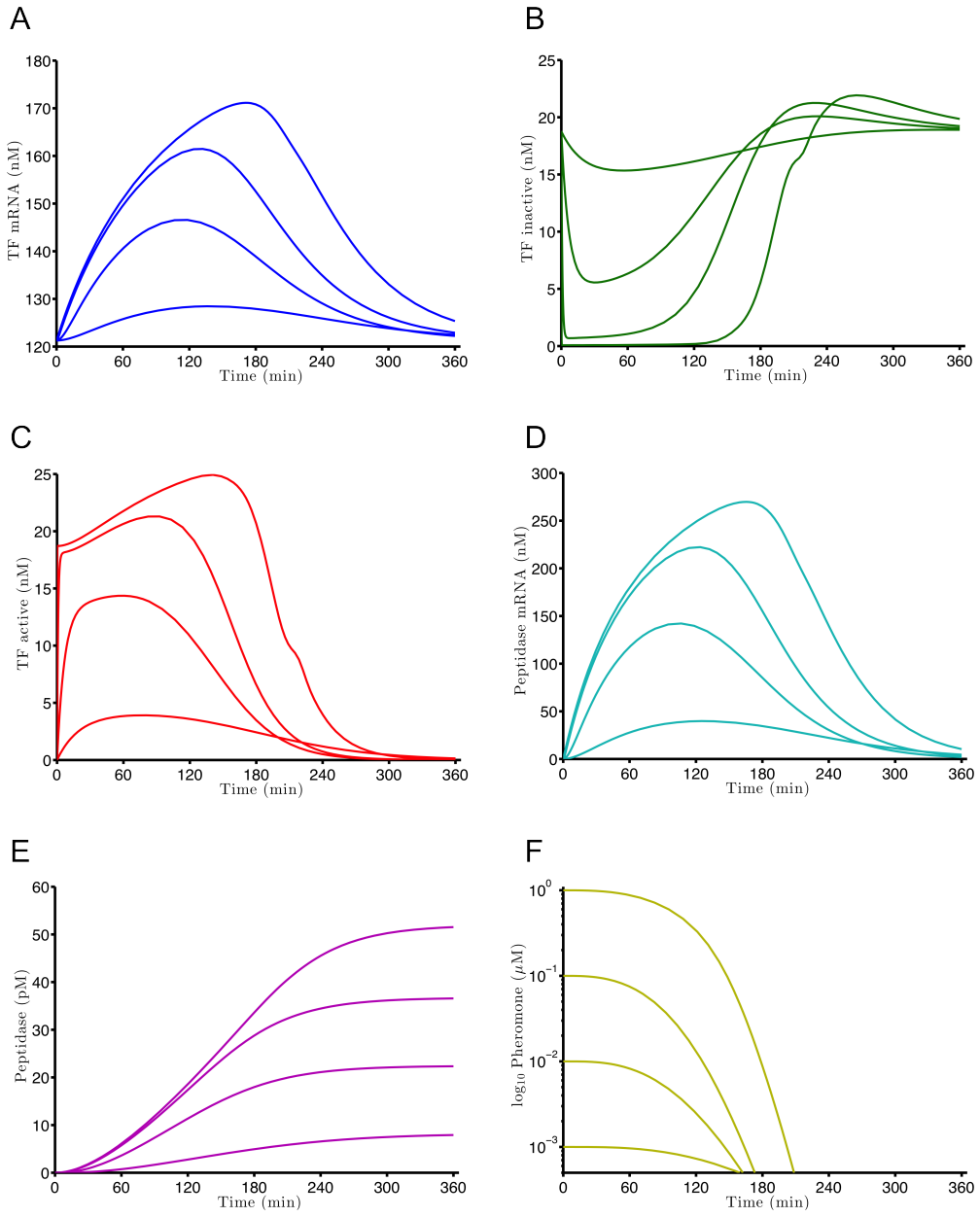


FIGURE 5.13: Simulation of model A2R with multiple doses of pheromone input (1, 10, 100, 1000 nM). (A) TF mRNA. (B) TF inactive. (C) TF active. (D) Peptidase mRNA. (E) Peptidase. (F) Pheromone.

different parts of the model. Setting the parameter k_c to zero removes the catalytic activity of Peptidase, analogous to a $\text{Sxa2}^{\text{S200A}}$ yeast strain, and corresponds to a model without negative feedback (Figure 5.14). Since there is no Peptidase activity in this variant, the strength of the pheromone input remains constant throughout the simulation (Figure 5.14 F). The constant input translates into a sustained activation of the TF pool (Figure 5.14 C), which reaches a steady state

that depends on the pheromone dose; however, at doses of 100 nM pheromone or greater, the response becomes saturated and the separation between trajectories becomes much smaller than in the wild type counterpart. This suggests that the purpose of a negative feedback through Peptidase activity is not limited to signal termination, but may also play a role in allowing the cell to accurately discriminate between different levels of input strength.

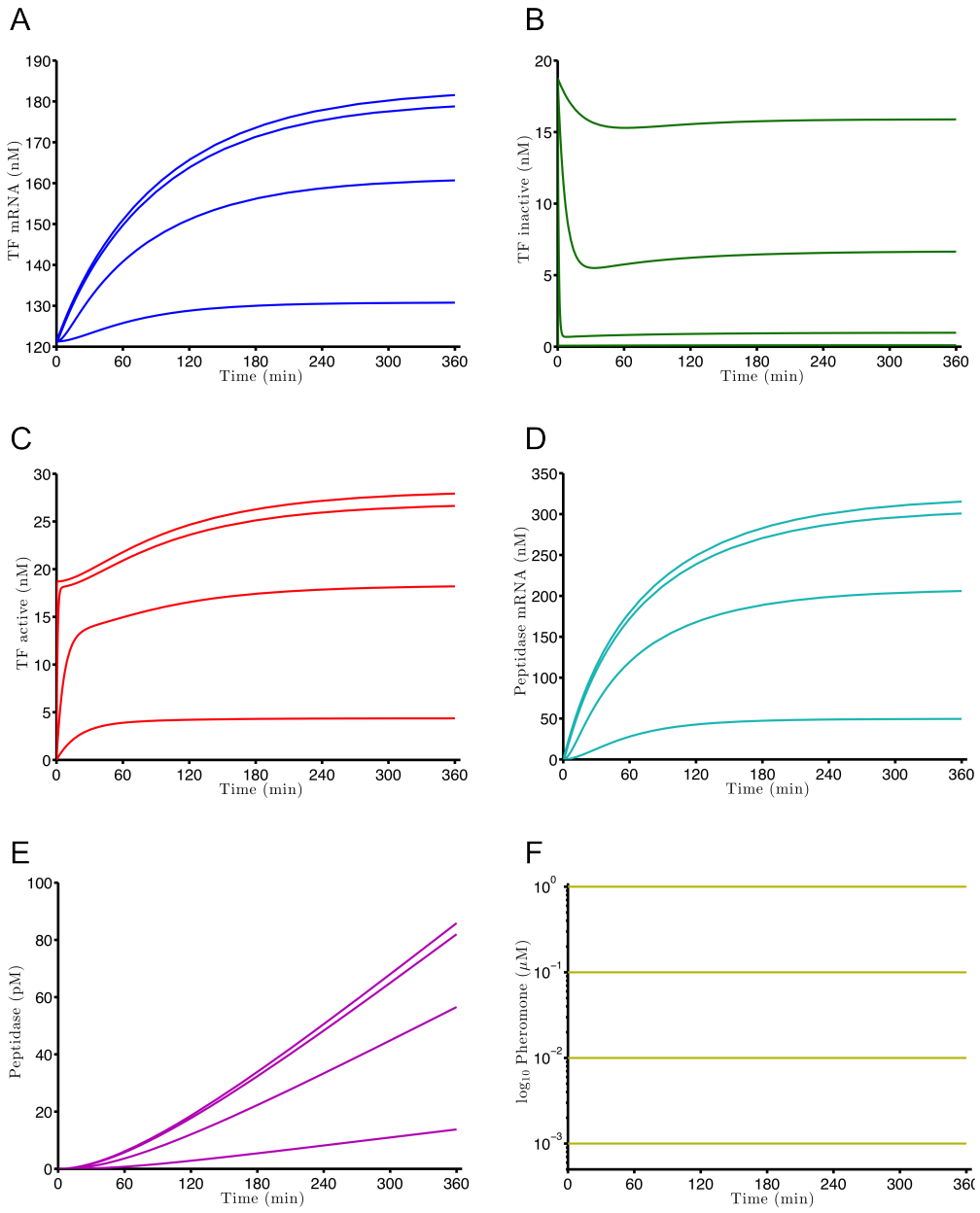


FIGURE 5.14: Simulation of model A2R without negative feedback using multiple doses of pheromone input (1, 10, 100, 1000 nM). (A) TF mRNA. (B) TF inactive. (C) TF active. (D) Peptidase mRNA. (E) Peptidase. (F) Pheromone.

It has long been recognised that in pheromone dose-response experiments using transcriptional reporters, the transition from the minimum to the maximum output response occurs over a very narrow range of concentrations (Didmon *et al.*, 2002; Das *et al.*, 2006; Smith *et al.*, 2009). Since most of the previous works in the study of pheromone signalling have been performed in *sxa2*$^{-}$ strains, it is possible that the perceived suddenness in the transition is a consequence of using *sxa2*$^{-}$ strains, and that wild type cells would display a more graded response. This prediction could be tested by performing quantitative gene expression time courses of *sxa2* in both wild type and Sxa2^{S200A} strains, with a selection of different pheromone doses guided by the model. Using existing data (Croft, 2012), it is possible to show that the model displays a qualitative agreement with a time course of a fluorescent reporter expressed from the *sxa2* locus in response to various doses of pheromone (Figure 5.15). However, due to sensitivity limitations in these assays, the experiments must be conducted over 12 hrs to achieve reliable measurements, which is much longer than the intended scope of the model presented here.

Setting the parameter v_1 to zero removes the positive feedback of TF towards its own expression. There are no analogous mutant yeast strains with this feature, so the model simulations are currently the only tool to explore this regulatory mechanism. The absence of positive feedback has a smaller impact on the model than the absence of the negative feedback (Figure 5.16). Since the level of TF mRNA remains constant in these simulations, the maximum level of TF active that can be produced is capped by the amount of TF inactive that was present in the cell before pheromone stimulation. This reduces the dynamic range available for TF activation compared to the wild type case (Figure 5.13 and 5.16), and could also hinder the ability of the cell to distinguish differences between certain pheromone inputs. In addition, the expression of TF-regulated genes becomes dampened in the absence of positive feedback, and the effect is not limited to high pheromone doses (Figure 5.17). This reduction in expression would have a greater impact

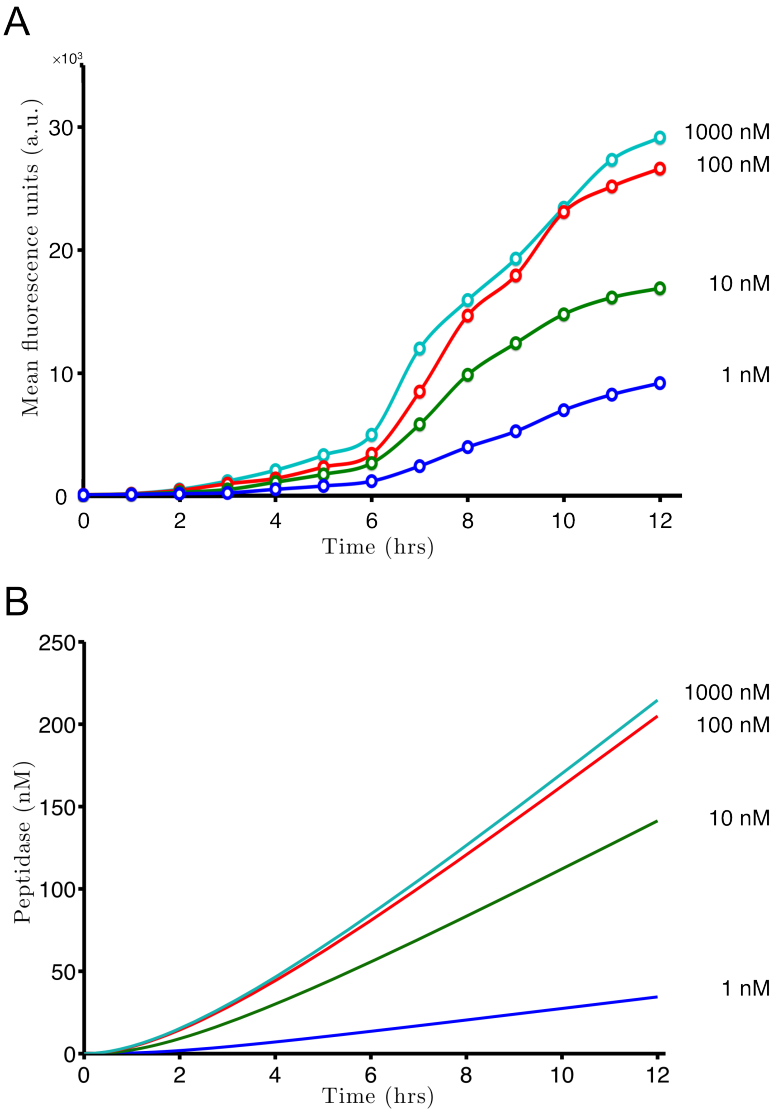


FIGURE 5.15: Qualitative agreement between fluorescent reporter time courses and model simulation. Data from Croft (2012) (A) where GFP is expressed from the *sxa2* locus in a *sxa2*$^-$ strain displays the same saturation as the model simulations without negative feedback (B) for pheromone doses above 100 nM.

upon lowly expressed genes, where even a small decrease in available transcription factor could result in a complete lack of expression (Del Vecchio, 2011).

A common idea in the study of cell fate decisions is that master regulators must cross a threshold in concentration to commit the cell to a particular programme of development (Purvis and Lahav, 2013; Li *et al.*, 2013). If such a threshold is necessary for *S. pombe* mating, either for Ste11 itself or for a Ste11-dependent gene, then the strength of the positive feedback is likely tuned so that the threshold is

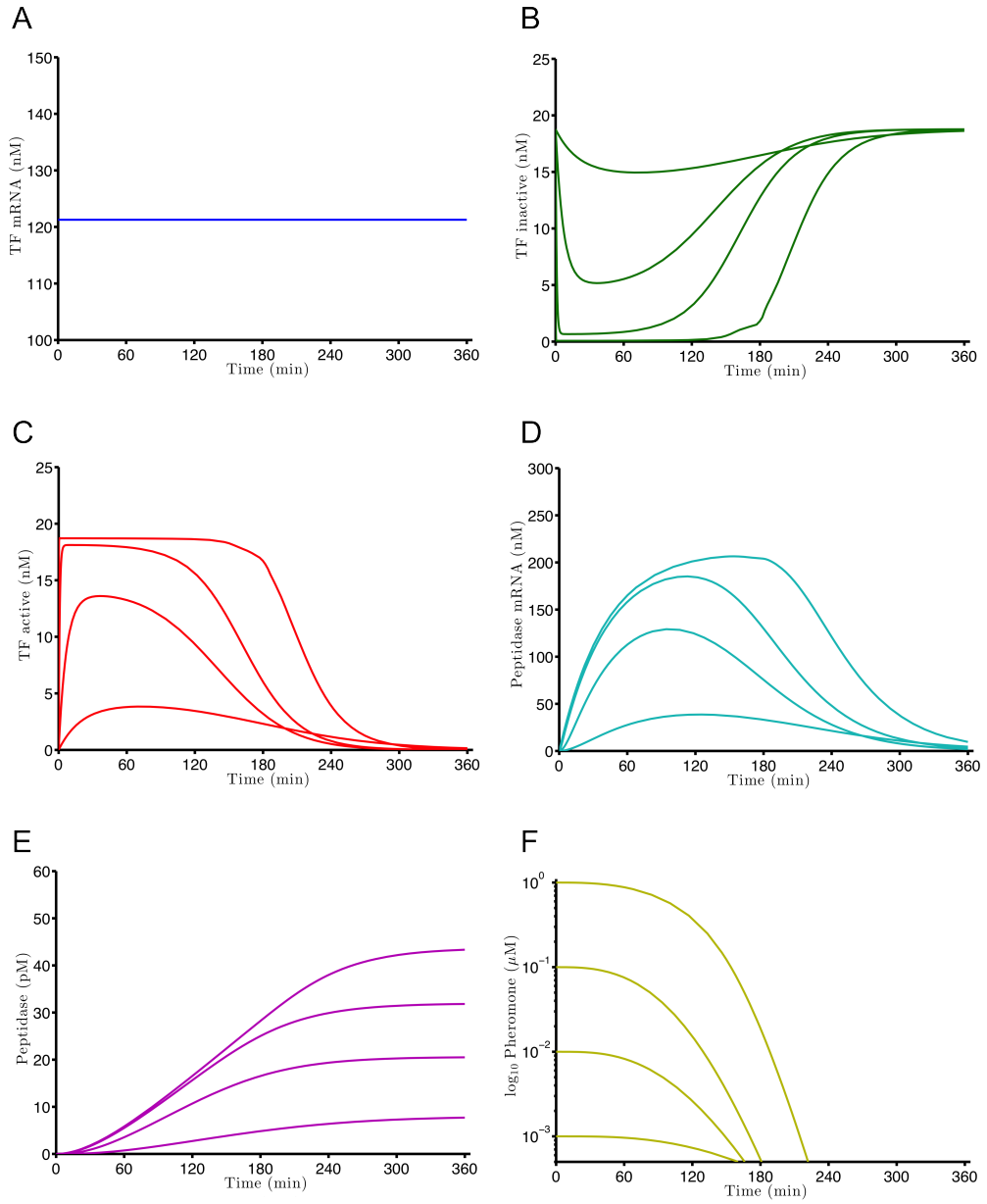


FIGURE 5.16: Simulation of model A2R without positive feedback using multiple doses of pheromone input (1, 10, 100, 1000 nM). (A) TF mRNA. (B) TF inactive. (C) TF active. (D) Peptidase mRNA. (E) Peptidase. (F) Pheromone.

only crossed in response to an appropriate pheromone input. A lack of positive feedback would mean that a higher pheromone dose would be required to cross that same threshold.

To investigate if a threshold mechanism could be inferred in this system, the trajectories of TF active were compared to a full dose-response curve of a transcriptional reporter strain, to assess if any differences in transcriptional output

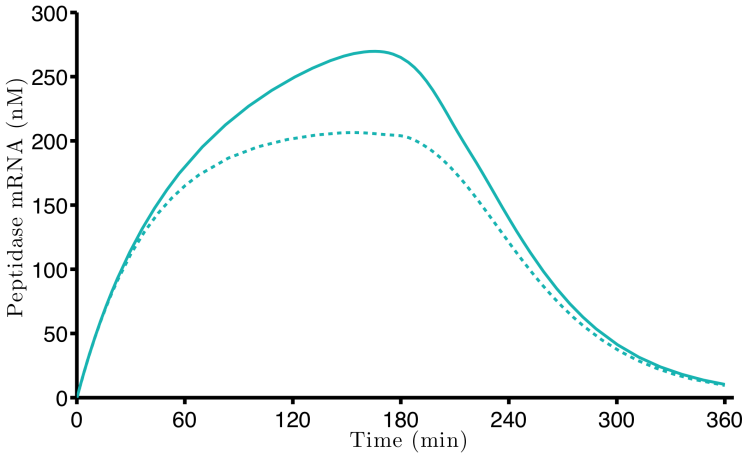


FIGURE 5.17: Lack of positive feedback dampens the expression of TF-regulated genes. Simulating model A2R with parameter v_1 set to zero (dashed) limits the extent of Peptidase mRNA production compared to a wild type simulation (solid).

could be matched to dose-dependent differences in the concentration of TF active (Figure 5.18). The most common null hypothesis is that transcriptional output should match the total amount of TF active generated during the time course (Coulon *et al.*, 2013); however, calculating the area under the curve for all trajectories showed that this relationship would lead to a graded linear response instead of the observed sigmoidal (Figure 5.18 B).

The idea of a concentration threshold could explain why an input dose of 10^{-8} M is capable of activating 75% of the pre-existing TF inactive molecules, but still produces an output that is indistinguishable from basal levels of activity (Figure 5.18). A natural concentration threshold could be set by requiring that all the TF pool be activated as seen for inputs of 10^{-7} M and above. This would explain why a small transcriptional output is observed for $10^{-7.5}$ M, as the corresponding TF active trajectory is able to briefly cross this threshold at around 60 min of simulation time (Figure 5.18). However, this idea alone would not explain why a pheromone dose of 10^{-7} M, which can activate of TF molecules as fast as higher doses and whose trajectory remains above the threshold for 2 hrs, is still incapable of producing a maximal output (Figure 5.18). Since these assays are performed on

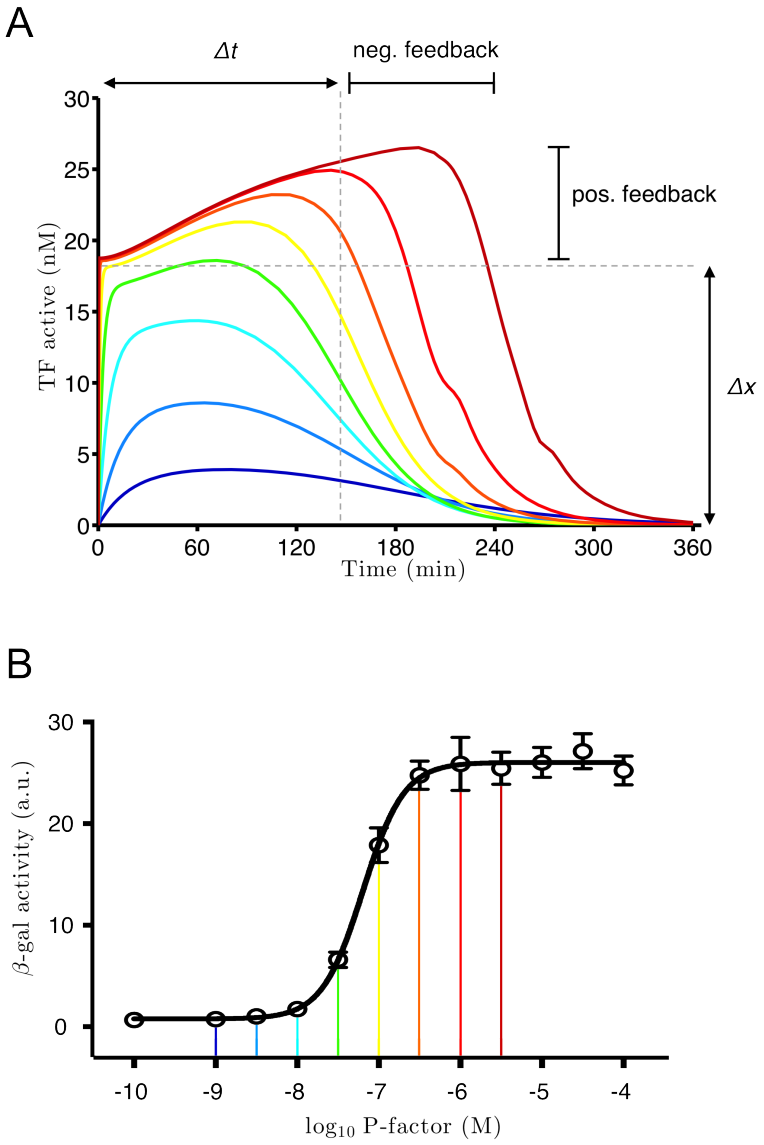


FIGURE 5.18: Dose and duration thresholds for commitment to mating. (A) Simulation of wild type model with various pheromone doses. The postulated concentration threshold Δx, and time threshold Δt are controlled by the strength of the positive and negative feedback loops respectively. (B) The yeast strain JY544 described by Didmon *et al.* (2002) harbouring a *lacZ* gene at the *sxa2* locus, was used to construct a dose-response curve to various concentrations of pheromone following the protocol developed by the authors. Coloured lines indicate the matching between the model simulations and the dose-response data.

a population of cells, it is possible that the sigmoidal dose-response relationship is an artifact of the experiment, masking an underlying all-or-nothing response at the single cell level, where the 10^{-7} M dose would induce most cells to produce a maximal output, but with a small unresponsive population giving the appearance

of a sub-maximal overall response.

The possible population effect can be shown not to exist through single cell image analysis of a fluorescent reporter strain, which shows that transcriptional output is not binary at the single cell level, and that for a 10^{-7} M dose most cells do not exhibit a maximal output response (Appendix E). This confirms the earlier suggestion that a concentration threshold is not sufficient as the sole explanation for the observed dose-response behaviour.

A possible alternative is that there are two thresholds that must be satisfied to produce a maximal response, with a time threshold delimiting the amount of time that TF active must spend above the concentration threshold in order to elicit a full transcriptional response. The two thresholds would divide the trajectories of TF active into four quadrants, where pheromone concentrations that produce a full transcriptional response have trajectories that are able to reach the upper right quadrant, while pheromone concentrations that only produce a basal response have trajectories whose initial activation is confined to the bottom left quadrant. Pheromone concentrations that are close to the EC50 would allow TF active to cross the concentration threshold, granting them intermediate levels of output, but would not be able to sustain the activation levels for long enough to cross the time threshold necessary for a full response (Figure 5.18).

Scanning the values of the parameters v_1 and k_c across a range from -50% to $+50\%$ of their nominal estimated values, showed that the strength of the positive feedback parameter determines the maximum level of TF active concentration that can be reached, without affecting the time at which the trajectory falls below a given concentration threshold, while the strength of the negative feedback determines the amount of time that a trajectory can stay at its maximum value, but without affecting the level of that maximum (Figure 5.18 A). This indicates

that the two feedback loops are independent, and that each serves to specify a particular threshold.

The requirement of a time threshold for full pathway activation is consistent with the known expression pattern of the mating transcriptional network, where there are distinct waves of Ste11-dependent gene expression, with some of the late genes requiring both Ste11 and early expressed genes to fully induce their promoters (Mata and Bähler, 2006; Mata *et al.*, 2007; Harigaya and Yamamoto, 2007). A drop in Ste11 activation levels before the final wave of genes has been expressed could then be a mechanism to abort the whole process.

The idea of a dose-to-duration encoding in the dynamics of regulatory proteins has been suggested before, as it allows distinguishable outputs to be obtained despite a saturating input, and has been demonstrated experimentally to occur in some signalling pathways (Behar *et al.*, 2008; Hao and O'Shea, 2012); however, previous works have only explained these dynamics through models with generic feedback loops which are not explicitly mapped to specific biochemical processes. Here, a plausible setting has been presented where a time and concentration threshold mechanism is wired directly into the known structure of a gene regulatory network to regulate the output of a transcription factor.

5.3.1 Sensitivity analysis

Sensitivity analysis is a standard model analysis tool that can provide information about the relative importance of each reaction towards specifying the model behaviour (Zi, 2011). The analysis consists of quantifying the magnitude of changes experienced by the state variables in response to perturbations in the parameter values. Here, classical time-varying sensitivities were used, also known as forward sensitivities (Ingalls and Sauro, 2003).

Following the notation introduced in section 3.5.1, where $\boldsymbol{x}$ is the n-vector of states, and $\boldsymbol{p}$ is the q-vector of unknown parameters, the time-varying concentration sensitivity coefficients are defined to be the $n \times q$ elements of the matrix function $\mathbf{R}_{\boldsymbol{p}}^{\boldsymbol{x}}(\cdot)$ given by

$$\mathbf{R}_{\boldsymbol{p}}^{\boldsymbol{x}}(t) := \frac{\partial \boldsymbol{x}(t,\boldsymbol{p})}{\partial \boldsymbol{p}}\Big|_{\boldsymbol{p}=\boldsymbol{p}_0} = \lim_{\Delta \boldsymbol{p} \to 0} \frac{\boldsymbol{x}(t,\boldsymbol{p}_0 + \Delta \boldsymbol{p}) - \boldsymbol{x}(t,\boldsymbol{p}_0)}{\Delta \boldsymbol{p}} \text{ for all } t \geq 0. \quad (5.27)$$

where $\boldsymbol{p}_0$ includes an initial condition (Ingalls and Sauro, 2003). These sensitivity coefficients give a measure of the amount of change that would be experienced in the state trajectories, between the nominal unperturbed case and a single parameter perturbation at a specific time point. A positive-valued coefficient indicates that the change due to the perturbation would result in an increase of the state variable, while a negative-valued coefficient would result in a decrease, with no change occurring when a coefficient is zero. The sensitivity coefficients are calculated at each step during numerical integration of the state trajectories in a simulation experiment, thus tracking how the influence of a particular parameter changes through time for each state.

To allow a direct comparison between coefficients, the sensitivities were made dimensionless by the normalisation

$$\left(\frac{p_j}{x_i(t)}\right)\left(\frac{\partial x_i(t)}{\partial p_j}\right) \quad (5.28)$$

for the ith state, and jth parameter. The time-varying sensitivities were calculated for the model A2R with different starting pheromone concentrations; however, the same relationships were obtained in every case, with no significant differences between doses (Figure 5.19). A common pattern was observed in the sensitivities of opposing processes, where their sensitivities evolve as mirror images of each other along the time axis, for example, the basal TF mRNA transcription rate k_b, and its degradation rate $k_{d_{m1}}$, appear to be coupled with opposing effects on the

trajectories of all species. The same pattern is exhibited by the pair of parameters controlling pheromone inactivation, k_c and k_{m_3}, and in some cases also by the translation and degradation of TF, k_{t_1} and k_{d_p} (Figure 5.19).

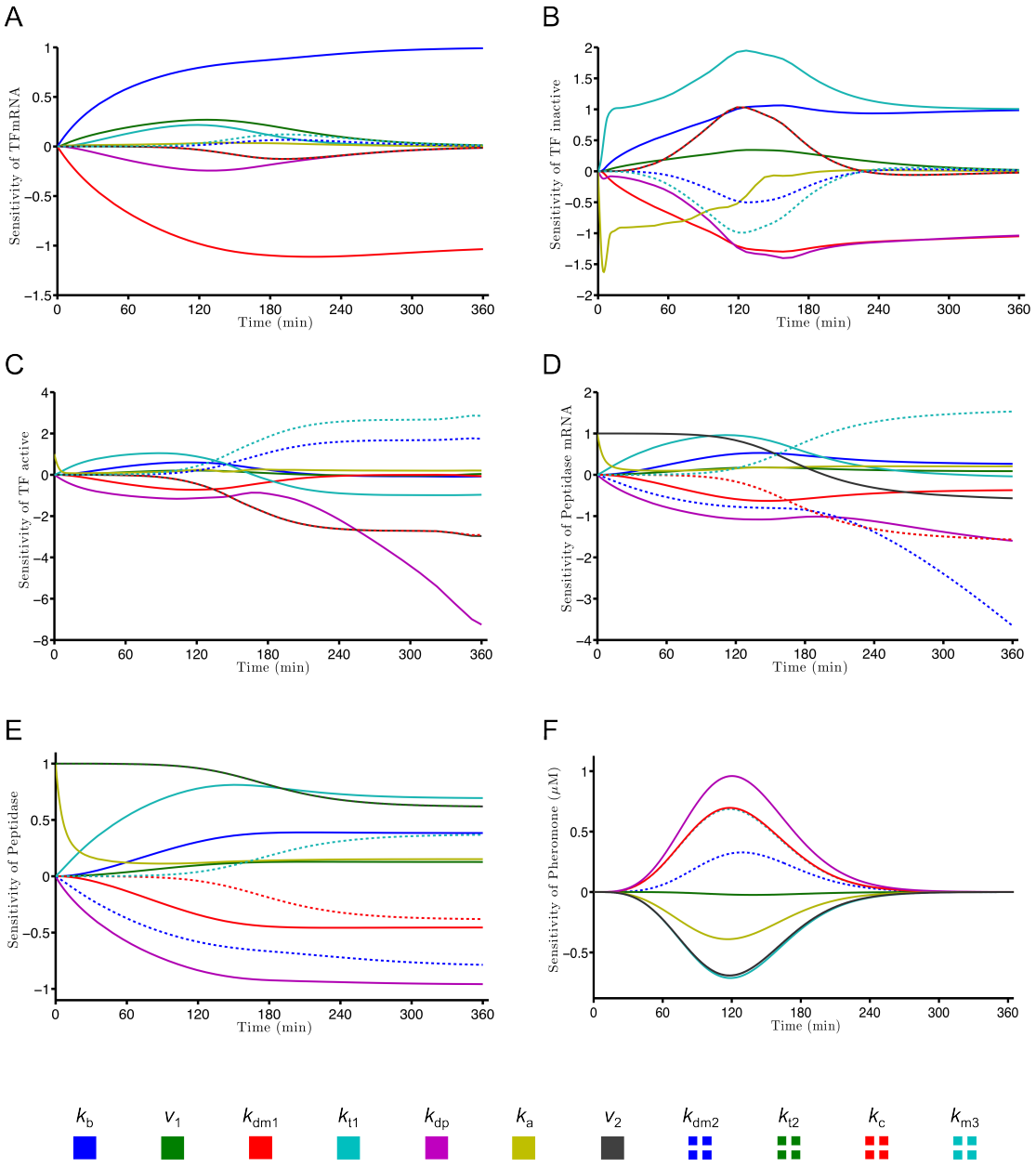


FIGURE 5.19: Time-varying sensitivities of model A2R.

Most parameters displayed a consistent positive or negative effect on species concentration throughout the entire simulation; however, two cases were found where the sensitivity of a parameter changed sign during the simulation. For TF active

(Figure 5.19 C), the translation rate k_{t_1} is initially promoting an increase in concentration, but as TF active approaches its peak level, the positive influence of k_{t_1} begins to diminish and eventually reverses its effect completely, working towards a decrease of TF active. A similar situation was observed for the effect of the transcription rate v_2 on the concentration of Peptidase mRNA (Figure 5.19 D). These effects were confirmed by simulating the model with different values of k_{t_1} or v_2, where an increase in parameter value caused the trajectories to reach a higher maximum, but was accompanied by a narrowing of the resulting parabola so that concentration levels dropped at a faster rate than before. The reversal occurs because a higher production rate in these species translates directly into a greater amount of Peptidase being produced at an earlier time, and to a corresponding faster clearance of the Pheromone stimulus.

The activation rate constant k_a, which was estimated to be the fastest process in the model (Table 5.5), only appears to be relevant during the initial rising phase of TF active, with little or no sensitivity displayed afterwards. The speed of activation seems to be directly related to its transient sensitivity, as simulating the model with successively slower activation rates results in the sensitivity of k_a remaining relevant for increasingly longer periods of time during the simulation.

As suggested before during inspection of the model simulations (Figure 5.13), the sensitivities of TF active and Peptidase mRNA confirm that their degradation rates, k_{d_p} and $k_{d_{m2}}$, begin to dominate over all other parameters after the pheromone input has been completely inactivated (Figure 5.19 C and D). To give a global picture comparing all the model sensitivities to each other, the time integral can be calculated for the sensitivity trajectories in Figure 5.19, to obtain a single value for each parameter-state combination (Figure 5.20). From this matrix it is readily apparent that the degradation rate of TF is the most sensitive reaction in the entire model. This parameter was fixed based on the estimated values obtained from model A1 (section 5.1); however, because of its high sensitivity it is

worth noting that rough estimates have been determined experimentally for this rate, with a reported half-life of 20 min for the Ste11 protein (Kjærulff *et al.*, 2007). This corresponds to a rate of 0.0346 s^{-1}, which is in very good agreement with the estimate obtained here of 0.0349 s^{-1}, and provides some confidence that the high sensitivity of k_{d_p} is a true feature of the pheromone signalling pathway.

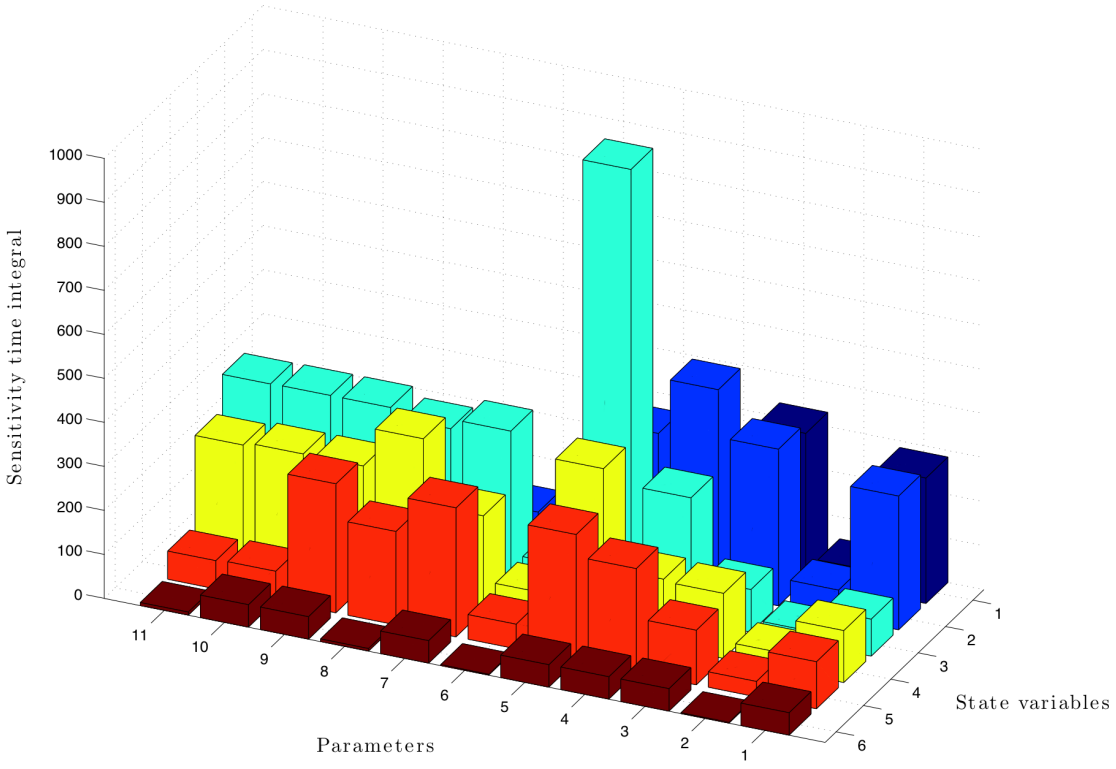


FIGURE 5.20: Sensitivity time integral of model A2R. State variables (1-6) are in the order: TF mRNA, TF inactive, TF active, Peptidase mRNA, Peptidase, and Pheromone. Parameters (1-11) are in the order: k_b, v_1, $k_{d_{m1}}$, k_{t_1}, k_{d_p}, k_a, v_2, $k_{d_{m2}}$, k_{t_2}, k_c, k_{m_3}.

A fast degradation rate for TF active is consistent with a role as a master regulator of a developmental fate decision, as it helps to prevent an unwanted pathway activation if the required inputs are not sustained for the required duration. Performing a parameter scan for k_{d_p}, showed that the speed of degradation can affect both activation thresholds postulated above, as a higher degradation rate limits both the maximum level of TF active that can be reached, as well as the time that can be spent above any concentration threshold.

The sensitivity of the positive feedback parameter v_1 appeared to be small compared to all other parameters (Figure 5.20). This was caused because the main factor specifying the TF active concentration was being predetermined by the high intracellular concentration of TF inactive prior to pheromone stimulation, and the influence of v_1 could only act to fine-tune the maximum level that was eventually reached. As mentioned during the model derivation stage (section 3.2), research into pheromone signalling is normally conducted using *cyr1*$^-$ strains to allow simplifying assumptions to be made about multiple cross-talking pathways that would otherwise interfere. Thus, the experimental system being modelled here does not represent a true wild type, where *ste11* would not have such a high basal level of expression. It is possible that in *cyr1*$^+$ strains the overall influence of positive feedback has a much larger impact upon signalling dynamics, since these strains are unable to fully derepress Ste11-independent *ste11* gene expression (Maeda *et al.*, 1990; Sugimoto *et al.*, 1991).

While ignoring any possible cross-talks, the influence of positive feedback in a background of lower basal *ste11* transcription can be explored by comparing simulations with and without positive feedback, in a model with a reduced value for parameter k_b, and with corresponding changes in the initial conditions (Figure 5.21). It was found that without the dominant influence of a high pre-existing concentration of TF inactive, the positive feedback could be a key factor in producing a timely response to mating cues, as the lack of positive feedback created a limitation in the amount of TF available for activation, which translated into a weak activation of downstream genes, and a subsequent protraction of the overall response time (Figure 5.21 right panels). The presence of positive feedback on the other hand allowed a robust expression of TF species to be maintained, and to produce a sharply defined response time (Figure 5.21 left panels). In addition, the roles of feedback in specifying time and concentration thresholds as discussed above were still valid under a reduced basal TF mRNA production regime.

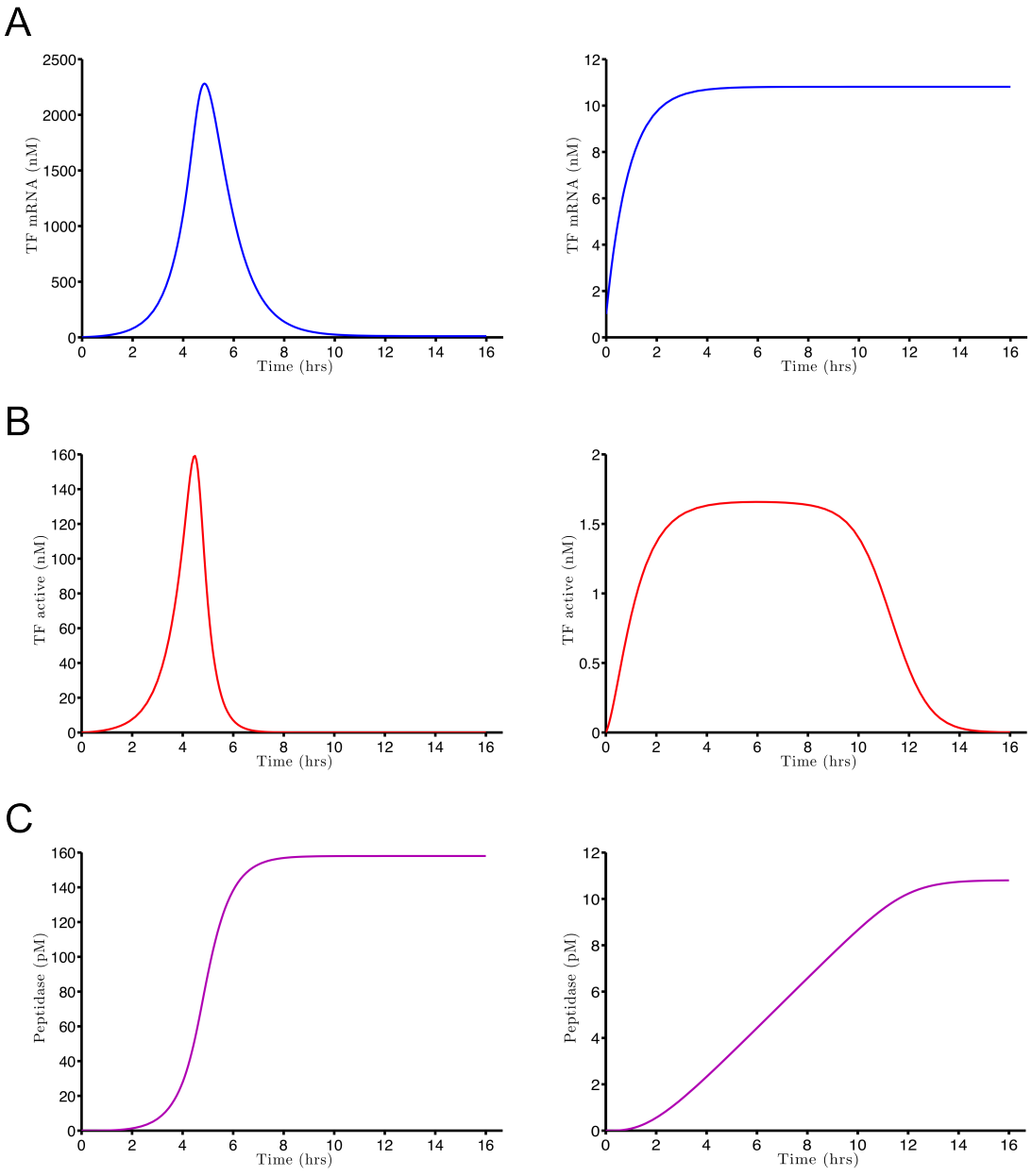


FIGURE 5.21: Importance of positive feedback in the absence of high basal TF mRNA production. The presence (left panels) and absence (right panels) of positive feedback was simulated in a model variant where the rate of basal TF mRNA was decreased ten-fold compared to the standard A2R model. (A) TF mRNA. (B) TF active. (C) Peptidase.

Chapter 6

Discussion and conclusions

In this thesis the roles of feedback control on pheromone signalling of *S. pombe* were investigated using a systems biology approach. To this end, a new mathematical model of pheromone signalling was developed to describe the dynamics of the master transcription factor Ste11, and the feedback loops that exist to regulate its function (Chapter 3). To encompass a full systems view of this process, the model includes aspects of cell signalling, gene expression, and enzymatic activity.

To gain a quantitative understanding of the speed and amplitude of model reactions, time course measurements were generated for as many model species as possible, focusing on technologies that allow absolute quantification (Chapter 4). The generated data sets were then used to estimate the rate constants in the model, and the accuracies of the estimated parameters were assessed to define the certainty of the model predictions (Chapter 5).

Comparing the model to data allowed the discrimination between several competing model variants to be made (sections 3.3 and 5.1.1), with further model analysis helping to elucidate the effects produced by feedback regulation in this system (section 5.3), as well as the relative importance of each process towards specifying the behaviour of the system (section 5.3.1).

Previous work to investigate this pathway has been mostly restricted to the analysis of individual gene functions, with little attention to the emergent behaviours that arise from the interaction of multiple system components. The work presented here represents a contribution towards an increased understanding of *S. pombe* pheromone signalling at a systems level, and of eukaryotic cell signalling in general.

6.1 The need for a new model of pheromone signalling

This project began with the intention of building upon an existing model of *S. pombe* pheromone signalling (Smith *et al.*, 2009), to study the influence of known feedback mechanisms on the dynamics of the system as a whole (Ladds *et al.*, 1996; Xue-Franzén *et al.*, 2006). The scope of the starting Smith *et al.* (2009) model focuses on the initiation of signal transduction upon pheromone detection by GPCR receptors at the plasma membrane, and the associated activation patterns of the transducing G protein Gpa1. Since the feedback loops that were of interest occur at a gene regulatory network level, the Smith *et al.* (2009) model would require a substantial extension to include downstream processes that would connect the events that occur at the plasma membrane with the initiation of pheromone-dependent gene expression, and to the feedback effects that arise from the induced gene products.

Although many of the steps that take place during *S. pombe* pheromone signalling are known (for a review see Otsubo and Yamamoto (2012)), initial work with the Smith *et al.* (2009) model quickly highlighted the fact that large gaps in the knowledge of this pathway still exist, and that the level of detail in the reactions of the Smith *et al.* (2009) model could not be extended to the entire pathway without

making assumptions that did not have a solid justification behind them. In fact, these gaps in knowledge had led to the proposal of a new mechanism of G protein function, a GTP-bound inactive state, in order for the Smith *et al.* (2009) model to successfully reproduce all the relevant experimental results; however, this inert Gα·GTP hypothesis represented a large deviation from established paradigms of G protein function (Wittinghofer and Vetter, 2011), which limited the effectiveness of the Smith *et al.* (2009) study to communicate its main goal: to present new data that could not be explained through the current understanding of the pathway, and the introduction of a mathematical model as a tool to locate the areas of ignorance that required further examination.

To arrive at the inert Gα·GTP hypothesis, the known relationships between molecular species involved in the G protein cycle were translated into a mathematical model, and the structure of the resulting network was then manipulated until the desired results were replicated. The main achievement of this process was to discover a network motif that could explain the input-output patterns in the data; however, there was a lack of consideration of whether the labels assigned to the state variables in the model still had a direct mapping to specific molecular species after the network rearrangement. Since the resulting network motif required that one round of G protein activation produced a single round of effector activation, the model was interpreted literally as requiring G proteins to become inert after activating a single effector. This situation illustrates the need for interdisciplinary experts that can bridge the communication gap between experimentalists and modellers in systems biology collaborations.

In section 3.1.1 of this thesis, it was shown that the motif structure discovered by Smith *et al.* (2009) could be reinterpreted in terms of accepted G protein functions, by postulating a plausible protein-protein interaction that would result in the required rate-limiting behaviour from a slow complex dissociation. The main limitation of the inert Gα·GTP hypothesis was that there does not seem

to be a way to falsify its mechanism, as the proposed inert Gα·GTP would be physically indistinguishable from active Gα·GTP. The rate-limiting complex dissociation mechanism is an equivalent hypothesis that does not require such a large re-evaluation of decades of G protein research.

Refining the Smith *et al.* (2009) hypothesis opened a new avenue for the investigation of G protein signalling in fission yeast, with straightforward falsifiable predictions such as the physical interaction between Gpa1, Ste6, and Ras1, which could be corroborated for example by co-immunoprecipitation; however, pursuing those experiments would have led this project away from the stated goal of studying feedback control at a systems level. Thus, the decision was made to develop a new model with an explicit focus on the species responsible for enforcing the feedback regulation.

6.2 Deriving a model of feedback control in pheromone signalling

The first step for building the model was to select the state variables to represent the feedback structure of the pathway. Since there was a lack of quantitative time course data in the literature, the model was chosen to be a minimal description of this system. The use of purpose-built mutants allowed several simplifying assumptions to be made, such as the decoupling that exists between nutritional monitoring and Ste11 regulation in the *cyr1*Δ strain (Valbuena and Moreno, 2010). Ste11 is constitutively expressed in these cells; however, its mere presence in the cell is not sufficient to exert its transcription factor functions, which can only begin after exposure to pheromone. This required that at least two possible states of Ste11 be considered, and since it was not clear whether the main determinant of Ste11 state was its subcellular localisation, or phosphorylation status, the two states

were broadly labelled as (TF) active and inactive (Qin *et al.*, 2003; Kjærulff *et al.*, 2005). Although there are many steps in-between pheromone detection by the receptors and Ste11 activation, the amount of pheromone sensed by cells is known to affect the rate of Ste11-mediated transcription (Didmon *et al.*, 2002), so this was reflected in the model by making the rate of Ste11 activation directly dependent on pheromone dose. To complete the model, the feedback loops of the pathway were included by incorporating a Ste11 enhancement of its own production (Sugimoto *et al.*, 1991; Kunitomo *et al.*, 2000), and the Ste11-dependent expression of *sxa2*, whose gene product is responsible for the enzymatic inactivation of P-factor pheromone (Imai and Yamamoto, 1994; Ladds *et al.*, 1996).

6.2.1 Gaps in knowledge give rise to a family of model variants

Deriving the model based on limited knowledge leaves open the possibility that alternative descriptions are better suited to model the system in question. In section 3.2.1 a number of possible model variants were defined based on the model details for which no optimal choice was clear. This included questions such as the level of detail that should be included, and how important were different aspects of the model to explain the actual signalling dynamics. The expectation was that the data generated at later stages of the modelling cycle would be used to resolve the ambiguities between the model variants.

6.2.2 Preliminary assessment of the new model

Before continuing with further efforts to develop the model, it was desirable to judge how well it could describe existing data, and to set the expectations of how much data would need to be generated to fully constrain the model dynamics.

Since most studies in this area to date have produced qualitative data, only one suitable data set could be found in the literature to which the model could be fit. The study by Ladds *et al.* (1996) performed a relative quantification time course of Sxa2 production in response to several doses of pheromone. Although the time consuming nature of the experiments used to acquire this data only allowed for single replicates to be obtained, it was expected that having information for multiple doses would make this data set highly informative for parameter estimation.

Fitting the model variants to the Ladds *et al.* (1996) data gave similar results in most cases, which highlighted the need for an objective way to discriminate between model fits. Typical approaches to model selection are broadly categorised as either selection criteria or statistical hypothesis testing, and although there are usually advantages and disadvantages that have to be weighed for each one, the large number of models that had to be compared simultaneously made the choice easier as hypothesis testing is usually restricted to pairwise comparisons (Motulsky and Christopoulos, 2004). Here the AIC score was used, as it is one of the most widely used criteria for model selection, and has been shown to compare favourably against other alternatives (Aho *et al.*, 2014). One of its advantages is that it not only gives information about which model is more likely, but it can also estimate how much more likely a model is compared to its competitors.

The use of AIC scores to compare model variants at this stage of the project, with only one data set available for parameter estimation, did not provide any conclusive evidence for or against particular model variants; however, it gave a clear indication that without additional data only the simplest alternatives should be considered. Nonetheless, in one particular case, the option between fixing Hill coefficients or leaving them as free parameters, a decision could be reached without resorting to AIC comparisons. A simulation analysis showed that deviations of the Hill coefficients away from a value of 1 resulted in model dynamics that were unambiguously incorrect, and that these effects could not be compensated by other

parameters (Figure 3.10). On the basis of these observations, the choice of model option was resolved for this variant.

6.2.3 Unidentifiable parameters and the possibility of a fully identifiable model

From the initial fits of models to the Ladds *et al.* (1996) data, it became clear that parameter non-identifiabilities were affecting the model, and several approaches were used to gauge the extent of dependency between the model parameters. A cross-correlation calculation showed which pairs of parameters were affected the most (Figure 3.12), and the resulting relationships were visualised using a Monte Carlo approach (Figure 3.13). This revealed several striking relationships where any parameter value was seemingly allowed, pointing towards the presence of structural non-identifiabilities. Other relationships showed that despite being strongly correlated, the parameter values were still confined to a small region of parameter space, which suggested that confidence intervals could be calculated to show how well determined each parameter was.

In principle, it is possible to compute confidence intervals through the pseudo-volumes of the data clouds in the Monte Carlo analysis (Balsa-Canto *et al.*, 2008); however, this analysis is computationally very expensive, so a more efficient alternative was sought. A literature survey revealed PLE analysis as a viable option, which in addition to being reported as robust for use with small or noisy samples, also had the benefit of having a readily available free implementation to perform the calculations (Maiwald and Timmer, 2008; Raue *et al.*, 2009, 2015).

The results of the PLE analysis were consistent with the results obtained through the other methods, but it also illustrated that they complemented each other, as they each provided a different perspective on the non-identifiabilities. Most parameters exhibited some type of non-identifiability, either structural or practical,

with only two parameters having finite confidence intervals in both directions (Figure 3.14). These results then prompted the question of whether a fully identifiable model was possible, and if so which measurements would be required to accomplish it. To answer this question definitively a structural identifiability analysis was performed using the STAUS method (Evans *et al.*, 2002). Although it was shown that full identifiability was possible, repeating the analysis with many different observation functions suggested that all model species had to be measured, at least in some combination, to resolve all the non-identifiabilities in the model (section 3.5).

6.3 Generating necessary data sets to constrain the model

6.3.1 Tracking pheromone inactivation by Sxa2

To measure the decline in pheromone concentration due to Sxa2 produced by cells, a new assay was developed that could be performed using basic laboratory equipment. Although in principle a more expensive solution was possible in the form of reversed-phase HPLC, the purchase of the necessary technologies did not seem cost effective as they would have been used exclusively for this experiment, with seemingly no further benefit to other projects in the research group. It was also not possible to find any such equipment within the university that could be used for this purpose.

The new assay consisted in sampling the media of P-factor treated cells and inactivating Sxa2 through heat treatment, followed with the measurement of the remaining pheromone concentration by recording the pheromone-induced cell volume increase of a *sxa2*Δ strain added to the test sample (Figure 4.1). Removing

the *sxa2*$^+$ cells being tested was necessary to ensure that no Sxa2 activity was carried over to further stages of the assay, but also so that they did not skew the measurements of cell volume. Since the cell volume measurements were performed on a population of cells, a number of measures of central tendency could be used to characterise the changes in the population. The measure that provided the most consistent results was the median, as it is robust to outliers and the distribution of the measured volumes normally displays heavy tails. To convert the cell volumes into P-factor concentrations, a standard curve was prepared through identical conditions as the samples, and the fitting of a dose-response equation to the standard curve allowed the interpolation of unknown concentrations from the cell volumes.

In previous years, brief attempts had been made by other people to measure the remaining concentration of pheromone in their samples with mixed results (Graham Ladds, personal communication). In the assay introduced here, several pitfalls were discovered during its development that required optimising the conditions of the protocol. The assay required the transfer of media containing pheromone between tubes after removing the *sxa2*$^+$ cells; however, the loss of pheromone due to adsorption to the tube walls was substantial enough to make the assay useless if not addressed properly (Figure 4.2). Interestingly, tubes manufactured for the express purpose of limiting protein loss due to adsorption were not effective for use with P-factor. The only solution that provided consistent results was to saturate the tube walls with a coating of purified BSA. Another confounding factor in the assay, which is usually not encountered in pheromone experiments, was the adverse effects caused by exposing cells to methanol if present by more than 1/100 of the culture volume.

The measurements obtained in this assay displayed very little variability between technical replicates; however, this reflects in part the limits of resolution capable by the cell counter used, as it was observed that the reported values fell into

discrete categories, instead of along a continuous scale, which implies that the equipment would round off measurements within a given interval to a single value. Nonetheless, the size of the intervals that could be resolved appeared not to distort the overall trends observed in the time course experiments.

From this assay a new data set was generated that reported the absolute quantification of changes in pheromone concentration in response to pheromone-stimulated Sxa2 production. The assay represents a cheap new addition to the tools available for studying pheromone signalling, especially with the use of *sxa2*$^{+}$ strains which have typically been ignored due to a lack of tools for their study.

6.3.2 Fixing the scale of Sxa2 concentration

The preliminary analysis performed on the model illustrated the drawbacks of using relative quantification data for parameter estimation, because it requires the introduction of additional parameters that will add to the non-identifiabilities of the model structure. The scaling parameter S_{pep}, which related the Sxa2 activities in the Ladds *et al.* (1996) data to the corresponding Peptidase variable, had two significant correlations with other model parameters, and a completely flat PL as a result of its structural unidentifiability.

Providing the model with measurements of Sxa2 on an absolute scale would allow the value of S_{pep} to be fixed, and remove its influence on the model parameters. To generate these measurements, the information obtained from previous reports that attempted the quantification of Sxa2 proved invaluable (Ladds, 1998; Ladds and Davey, 2000), as they had established the need for concentrating the medium samples in order for the concentration of Sxa2 to be within detectable limits. This information made these the most straightforward experiments of the whole project. It was unexpected however, to find that adding a GFP tag would remove the ability of Sxa2 to degrade P-factor, although it was the first time to add

a fusion tag bigger than just a few residues in length. It is known that Sxa2 goes through extensive proteolytic processing during its traffic to the extracellular space, so it is likely that the tag prevents the full maturation of Sxa2.

6.3.3 Transcriptional dynamics of *sxa2* and *ste11*

Quantifying the expression patterns of *sxa2* and *ste11* in response to pheromone was a natural step in the search for additional measurements to complement the protein data available for Sxa2, since gene expression analysis by qPCR is one of the most common measurements performed in current biological research, and as such the necessary equipment to perform these experiments is usually readily available.

The selection of reference genes for the normalisation of samples was facilitated by existing data sets of whole genome expression changes across a wide variety of perturbations, which allowed the list of ideal candidates to be narrowed down to a select few before having to perform any experimental validation. Since the efficiency of amplification plays a large role in determining the accuracy of qPCR measurements, great care was taken to design optimal primer pairs, with additional consideration of the amplicon that would be produced.

One of the most challenging aspects of generating this data set was the isolation of total RNA from yeast cells. Extracting RNA is usually technically demanding, but is notoriously more laborious to perform on *S. pombe* cells due to their tough external cell wall. Fortunately, the research groups that had to optimise this technique for the whole genome experiments mentioned above, also established the widely accepted gold standard protocol for RNA isolation from *S. pombe* (Lyne *et al.*, 2003).

The measurements from qPCR are usually made on a relative scale, where the changes in gene expression are expressed as fold differences with respect to a sample that is chosen as the unit calibrator. To continue the trend of the absolute quantification data sets that were generated previously, it was desired to transform these data from relative to absolute values. This is possible in principle by generating a carefully constructed standard curve from in vitro transcribed RNA of the target of interest; however, in practice this option is rarely pursued since it is difficult to perform, and the accuracy of the standards decrease very rapidly due to instability (Collins *et al.*, 1995). Thus, it was serendipitous that the opportunity arose to quantify the existing time course cDNA samples by ddPCR. The measurements performed by ddPCR were not only superior by providing absolute quantification, but the robustness of the technique removed most of the technical variability that was observed with qPCR.

The time course gene expression profiling of *sxa2* and *ste11* by ddPCR provides the best characterisation of their transcriptional dynamics, both in accuracy and temporal resolution, that has been reported to date. Xue-Franzén *et al.* (2006) had previously reported a pheromone-dependent upregulation of *ste11* of ∼3-4 fold; however, their results were obtained from DNA microarrays which are subject to a greater amount of variability than both qPCR and ddPCR. The data presented here shows that at most a ∼1.5 fold upregulation occurs for *ste11* transcription in response to pheromone.

6.4 Using the model to increase knowledge and understanding

6.4.1 Model fitting and discarding model variants

Fitting the models to the data generated in chapter 4 was done first for the original model A1B1C1D1E1. The model showed good agreement with the data; however, some non-identifiabilities remained since it was not possible to secure measurements for all species as originally intended. Having a model that was not fully constrained by the data, the question of model discrimination was posed more rigorously through an indistinguishability analysis, to determine if the model could be distinguished from alternative variants with the measurements provided.

Since the gene expression data obtained for *ste11* suggested that only a small increase was taking place in response to pheromone, it was desired to establish if a model without positive feedback would provide a better explanation of the data. The indistinguishability analysis confirmed that the measurements given could be sufficient to make the discrimination, while comparing the AIC scores showed that the standard model including a positive feedback was indeed the superior option; however, the comparison was not fully satisfactory because the positive feedback was only affected by the data indirectly through other model reactions.

Measurements that directly affected the positive feedback were available from the expression data of *ste11*, so the analysis was repeated using the equivalent A2 model, which was necessary to allow the fitting of mRNA dynamics. This model would also be unidentifiable, but again allowed the distinction to be made between models with or without a positive feedback loop. The model including a positive feedback was found to be decidedly better suited to explain the data, reinforcing the previous result.

During the transition from a model with one-step process for gene expression to the model with a two-step process, the parameter for transcription factor degradation k_{d_p} was fixed based on the estimates obtained with the simpler model. This would decrease the degrees of freedom during the estimation procedure, and could alter the results that would be obtained otherwise; however, since the estimated value was in very good agreement with previous experimental reports (Kjærulff *et al.*, 2007), this increased the confidence that this rate constant had been well determined and could be fixed at the estimated value.

Further indistinguishability analyses were attempted to resolve the best option for each model variant; however, the results were inconclusive as the computational complexity escalated too quickly before a positive answer could be found. Nonetheless, there were still grounds to resolve one last option. The model variants of type B considered whether it was necessary to introduce a delay between pheromone detection and transcription factor activation, and one of the parameters that was found to be identifiable was the activation rate constant k_a. Based on the estimated value of k_a, and on the simulation analyses that would follow, this process occurs extremely fast, which allows the model including a time delay to be eliminated with confidence, even when the indistinguishability analysis was inconclusive.

6.4.2 Final model iteration and the measurements required for full identifiability

To determine the extent of uncertainty that remained in the model due to unidentifiable parameters, the parameter vectors along each PL were used to simulate the model and visualise the spread between the resulting trajectories. This uncertainty can be exploited to determine the most informative experiments that could

be performed to minimise the spread of the trajectories, and thus the uncertainty in the model.

Several experimental designs were tested but none managed to constrain the model enough to achieve full identifiability, even if all species could be measured. This suggested that the model was too complex given the information content provided by the noisy data. Since a reduction of noise in the data did not seem like a viable option, the discrepancy can be resolved through model reduction.

A reduced version of the A2 model (A2R), was shown to become fully identifiable with the addition of measurements of total Ste11, suggesting an immediate avenue for further progress along this line of investigation. Since the model reduction was performed based on insensitive parameters, the quality of the fits to data obtained previously were not affected, and model A2R became the final version of the model developed in this work.

6.4.3 Discovering roles for feedback control in *S. pombe* pheromone signalling

Although full identifiability of the model was not obtained with the available data sets, the remaining uncertainty in the model was completely localised to the concentration scale of transcription factor, which did not preclude the possibility of drawing conclusions based on other aspects of the model. To gain a sense of the nominal behaviour of the system, the model was simulated using a range of pheromone doses that cover a standard dose-response experiment in *S. pombe*. Comparing the normal model behaviour to simulations lacking either one of the feedback loops revealed roles for both of them in the discrimination of elevated doses of pheromone.

The proposed mechanism assumes the existence of thresholds that define the point at which the cell commits to the sexual development programme. A concentration threshold dictates the minimum amount of active transcription factor that must be present in the cell to fully launch the transcriptional network, while a time threshold is defined by the period of time over which full activation must be sustained for the commitment to become irreversible. Perturbation of the parameters responsible for feedback then shows that the strength of the feedback loops determines the dose of pheromone that is required to satisfy both thresholds, with each feedback having exclusive control over one of the thresholds. The positive feedback controlling the concentration threshold, while the negative feedback controls the time threshold. These roles for feedback control go beyond simply enhancing expression or terminating signalling, and can only be understood from a systems level perspective.

The existence of these thresholds in cell fate determination have both experimental as well as theoretical support. For example, simulations have shown that the number of downstream genes activated by a transcription factor determines the concentration it must have in order for all genes to be transcribed robustly (Del Vecchio *et al.*, 2008). Since Ste11 has at least 78 genes under its control (Mata and Bähler, 2006), this provides the basis to postulate a concentration threshold. Evidence for a time threshold comes from observing the need for Ste11 to remain active for the induction of both early and late genes in the network (Mata *et al.*, 2007; Harigaya and Yamamoto, 2007), and supports the idea that converting a pheromone input dose to the duration of Ste11 activation is one of the mechanisms employed by *S. pombe* for cellular decision making.

The idea of dose to duration encoding in cell signalling has been demonstrated experimentally from an input-output point of view (Behar *et al.*, 2008; Hao and O'Shea, 2012), but the explanation of how it arises in a specific pathway has not been shown. The model developed in this work provides a new tool to explore

this behaviour with direct correspondence to specific molecular species. In addition, the availability of strains without Sxa2 activity provides the opportunity for a number of model predictions to be tested immediately through similar measurements to those performed in this work, simply by using additional pheromone doses, for example the prediction of saturation in the transcription of *sxa2* above 10^{-7}M pheromone. Finally, the model can be used to explore the behaviour of pheromone signalling in scenarios that are not yet available experimentally, such as the importance of the positive feedback in the absence of high basal transcription factor production.

6.5 Conclusions

The work presented in this thesis has detailed a systems biology investigation of pheromone signalling in *S. pombe*, with the aim of developing a quantitative dynamical model that would allow the study of feedback control in this system. To develop the model required a careful delineation of the current understanding of the pheromone signalling pathway, and illustrated the difficulties that can be encountered before obtaining meaningful information from mathematical models. A new assay had to be developed to overcome equipment limitations, but can now be used as an additional tool to investigate pheromone signalling. In addition, the data sets that have been generated here provide the finest level of detail to date for these species, and will inform other studies in this field. Successive refinement of the model allowed competing hypotheses to be rejected along the way, and the final version of the model has allowed a greater understanding of the critical factors that control signalling behaviour, and has discovered roles for the coupled positive and negative feedback loops towards information processing in the cell. Most of the goals established in the project aims were fulfilled to a large extent; however, the lack of measurements for all model species, and the possibility of

validating model predictions leave many challenges ahead for the continuation of the modelling cycle in the study of pheromone signalling.

Appendix A

Uncertainty in trajectories of model A2B1C1D1E1

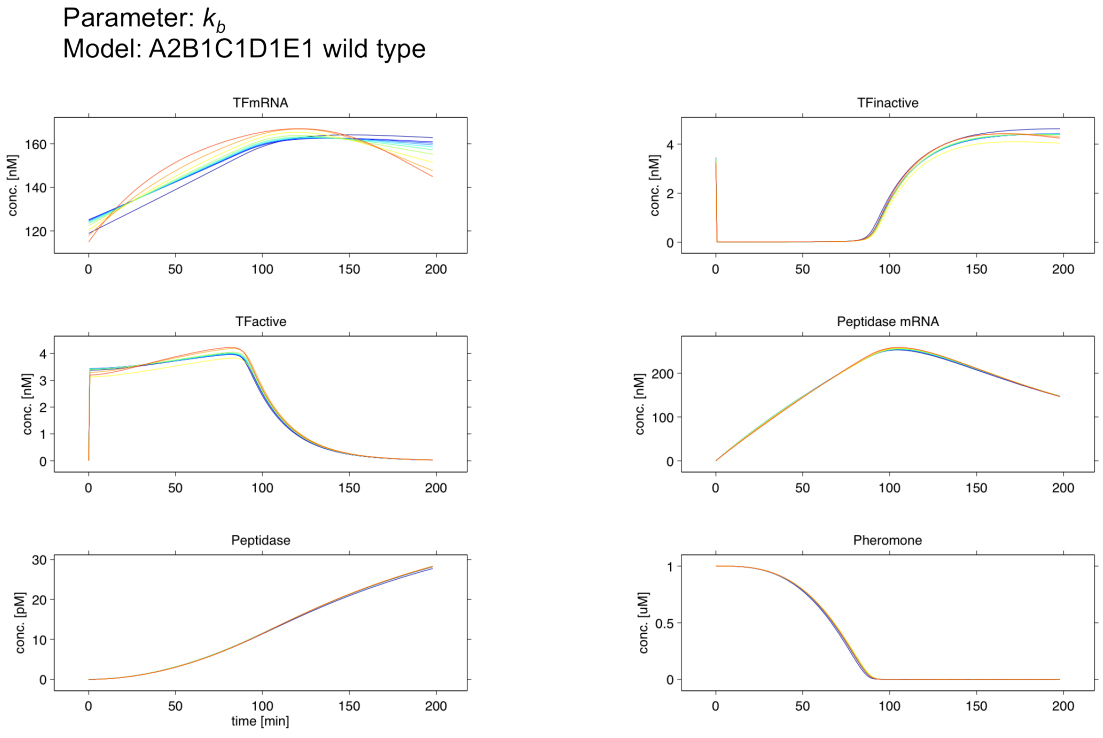


FIGURE A.1: Uncertainty in wild type model trajectories propagated from uncertainty in k_b.

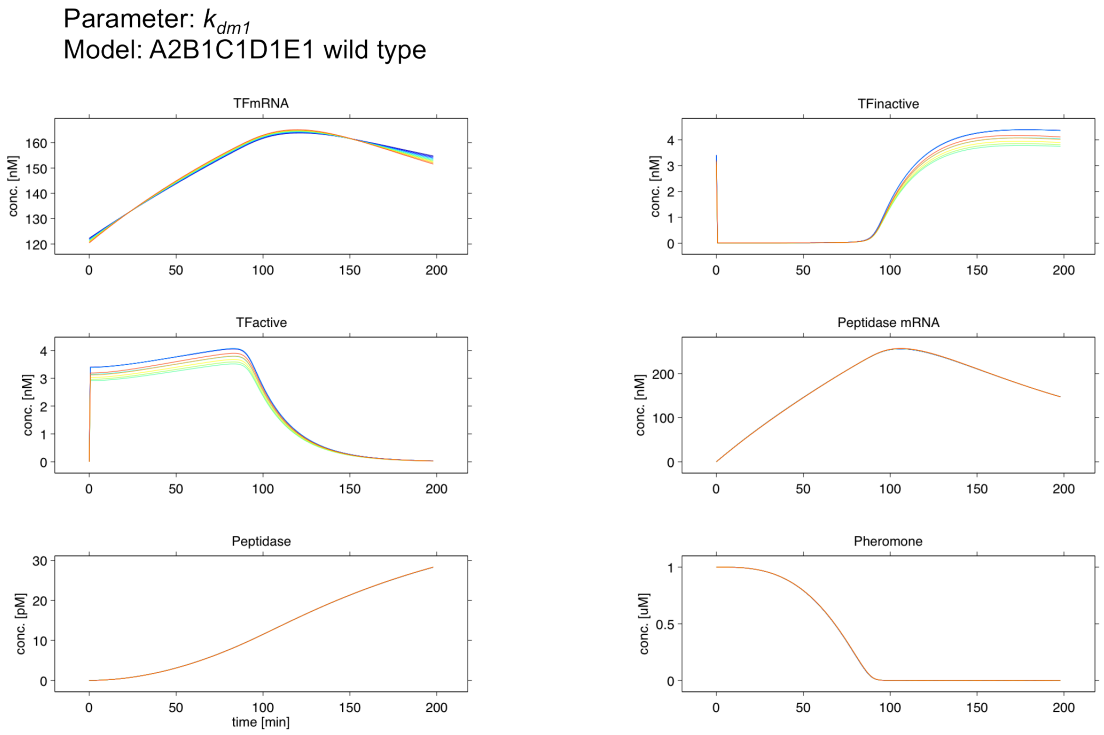


FIGURE A.2: Uncertainty in wild type model trajectories propagated from uncertainty in $k_{d_{m1}}$.

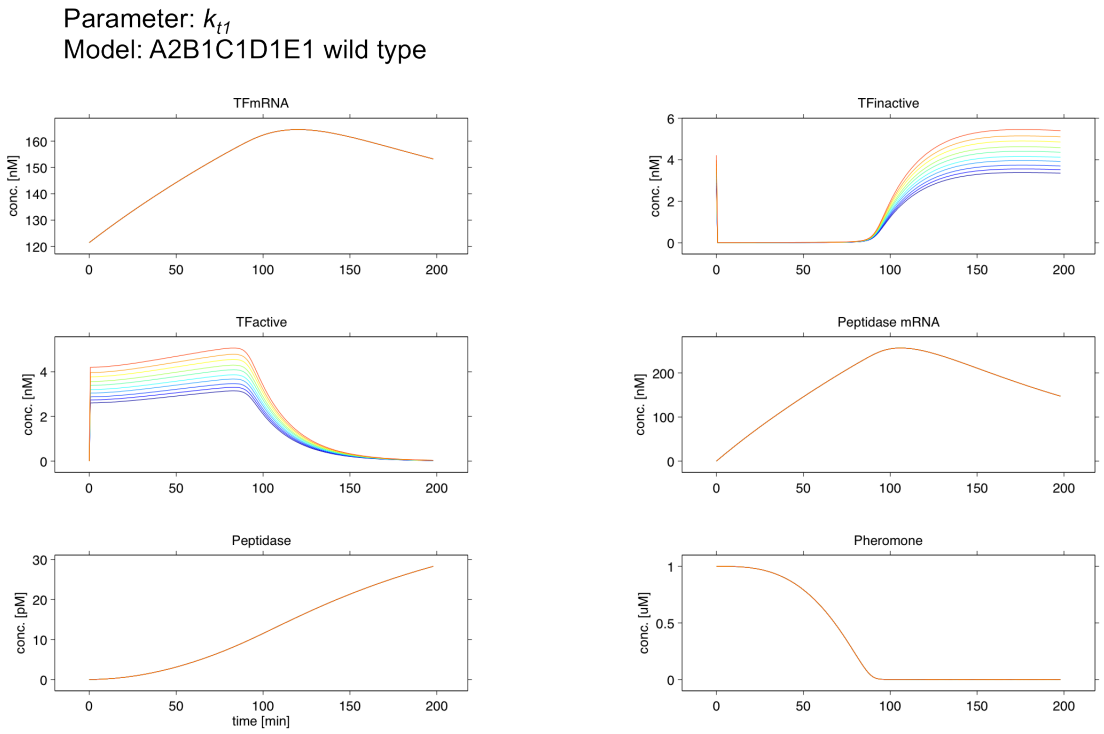


FIGURE A.3: Uncertainty in wild type model trajectories propagated from uncertainty in k_{t_1}.

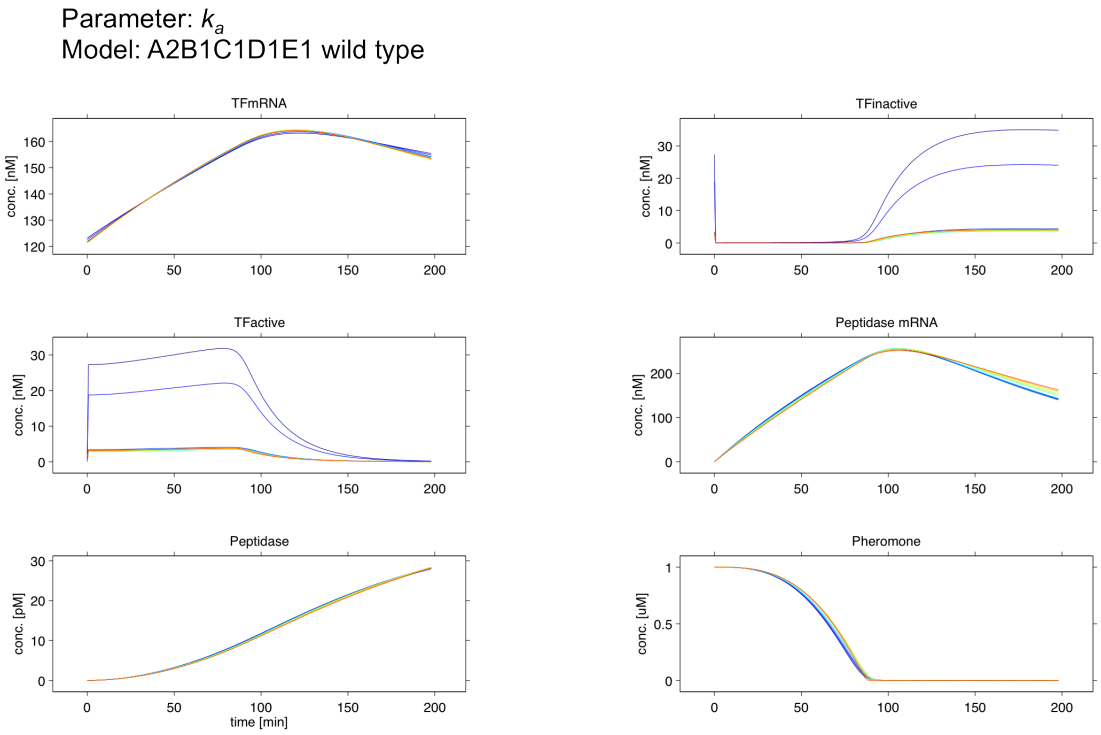


FIGURE A.4: Uncertainty in wild type model trajectories propagated from uncertainty in k_a.

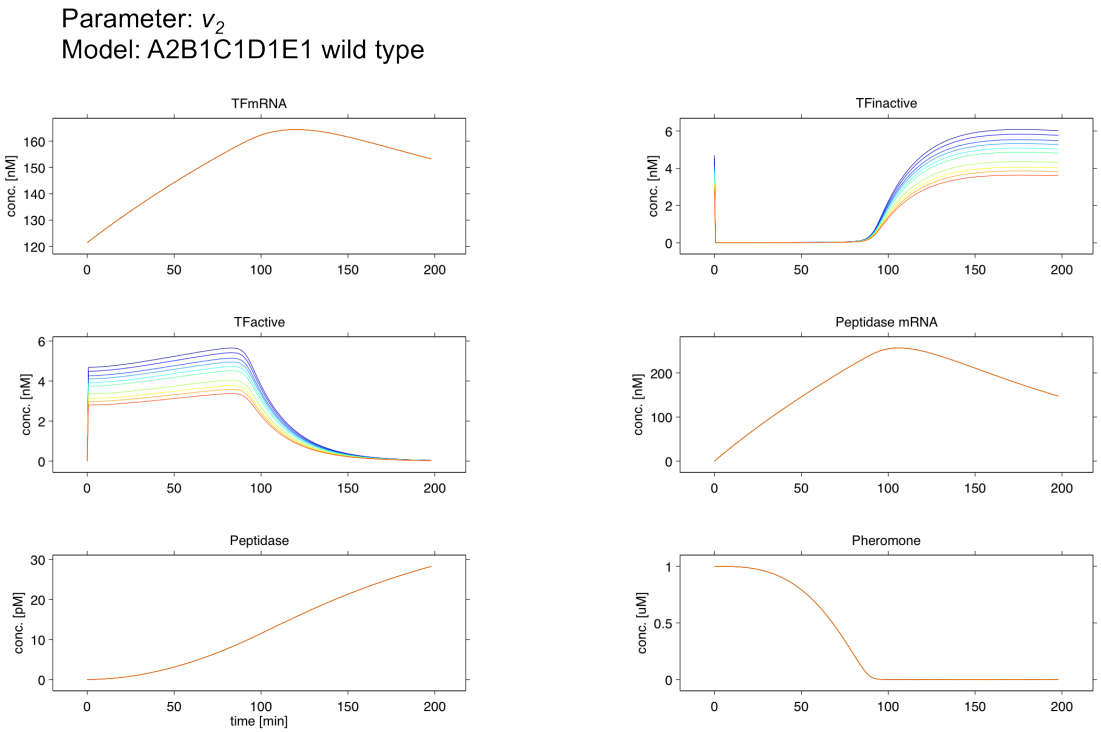


FIGURE A.5: Uncertainty in wild type model trajectories propagated from uncertainty in v_2.

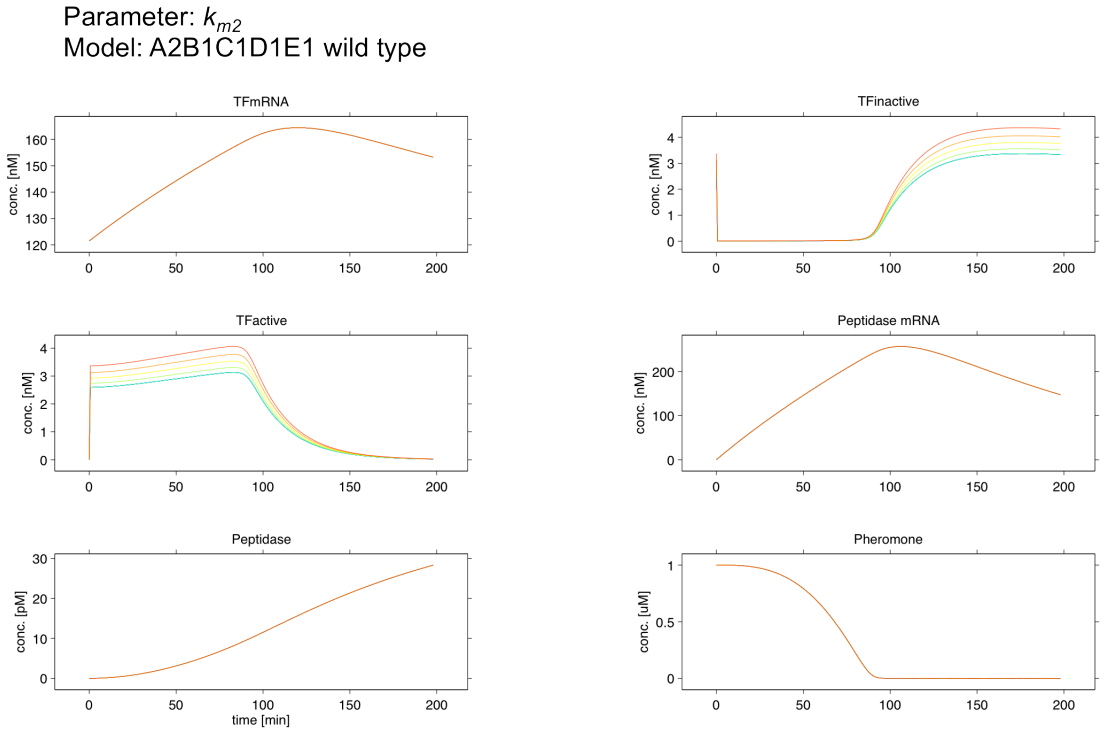


FIGURE A.6: Uncertainty in wild type model trajectories propagated from uncertainty in k_{m_2}.

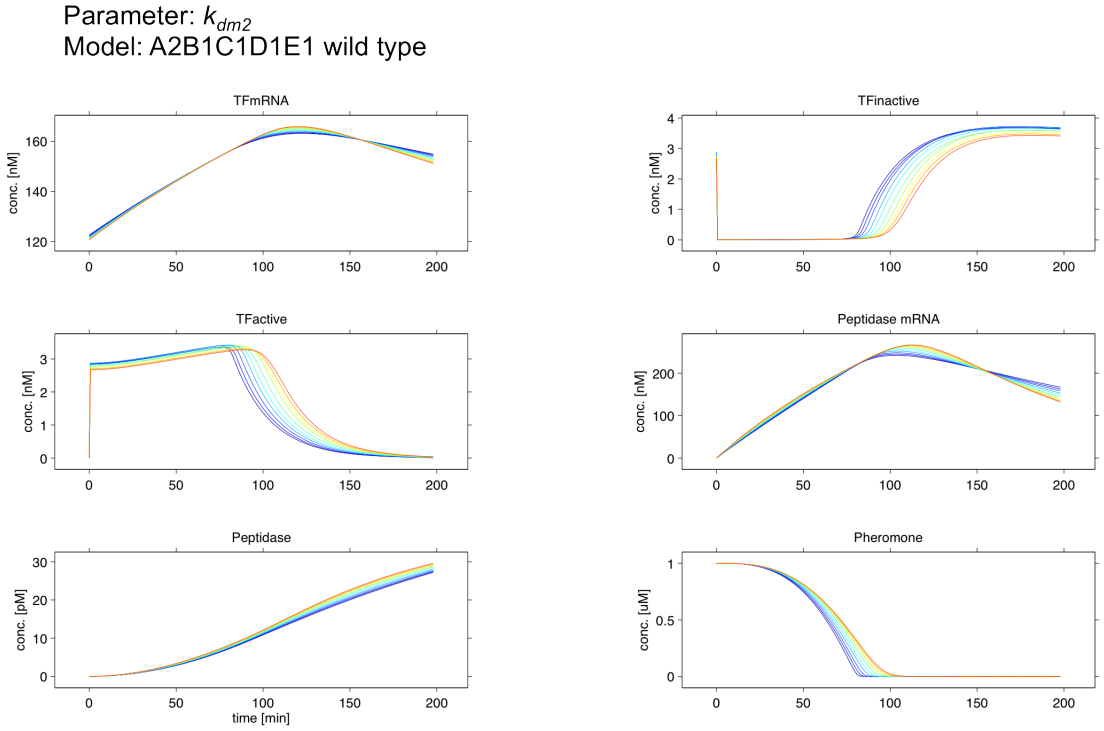


FIGURE A.7: Uncertainty in wild type model trajectories propagated from uncertainty in $k_{d_{m2}}$.

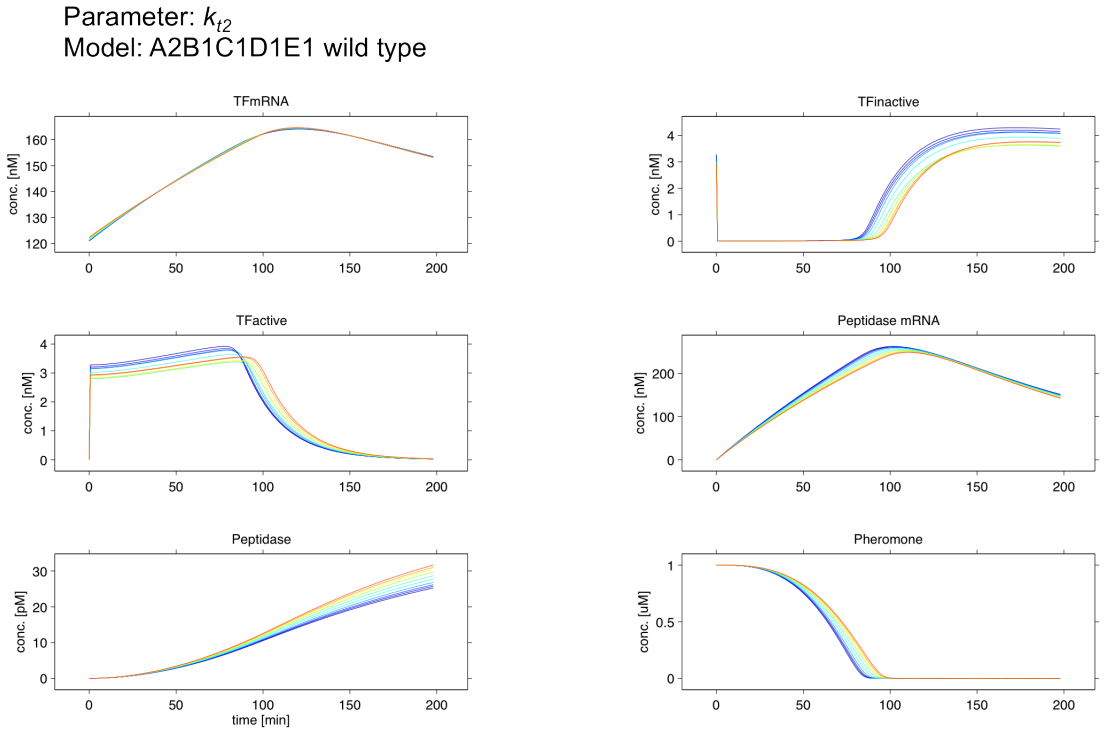


FIGURE A.8: Uncertainty in wild type model trajectories propagated from uncertainty in k_{t_2}.

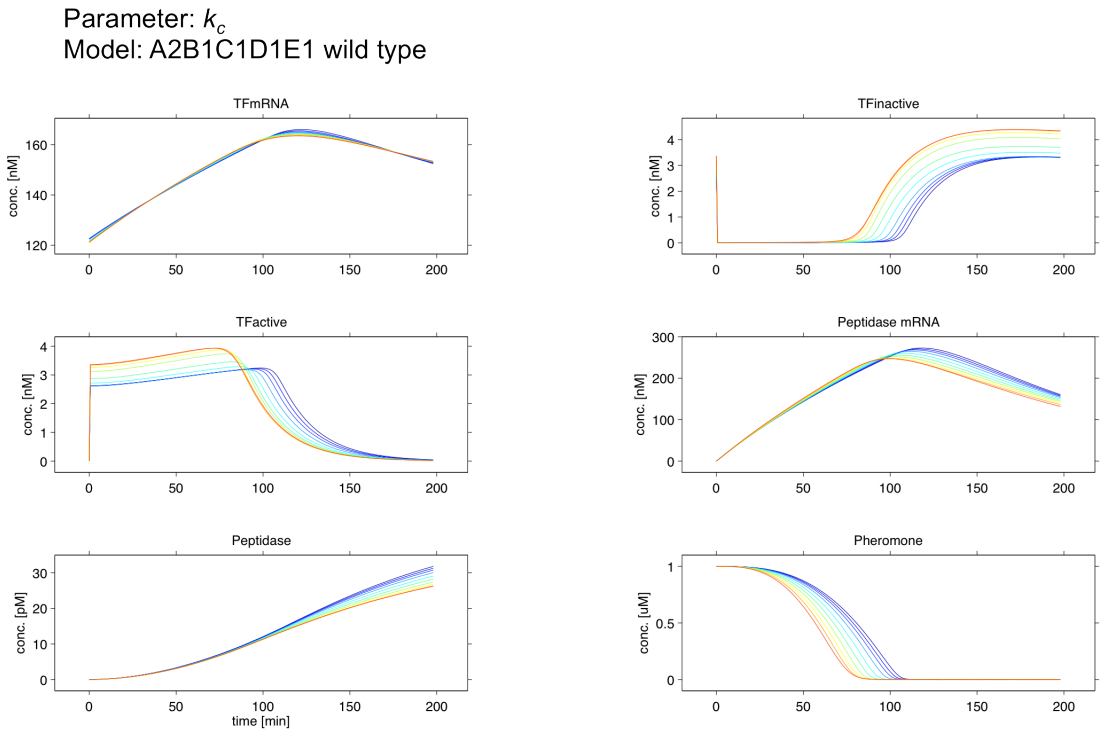


FIGURE A.9: Uncertainty in wild type model trajectories propagated from uncertainty in k_c.

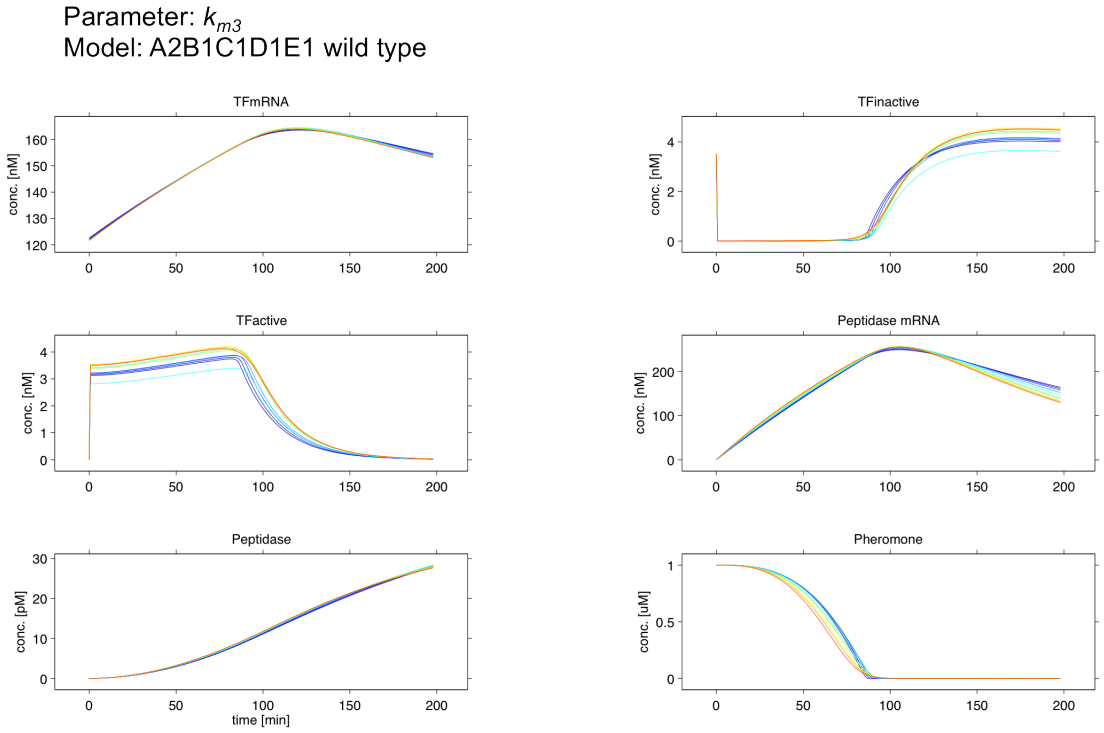


FIGURE A.10: Uncertainty in wild type model trajectories propagated from uncertainty in k_{m_3}.

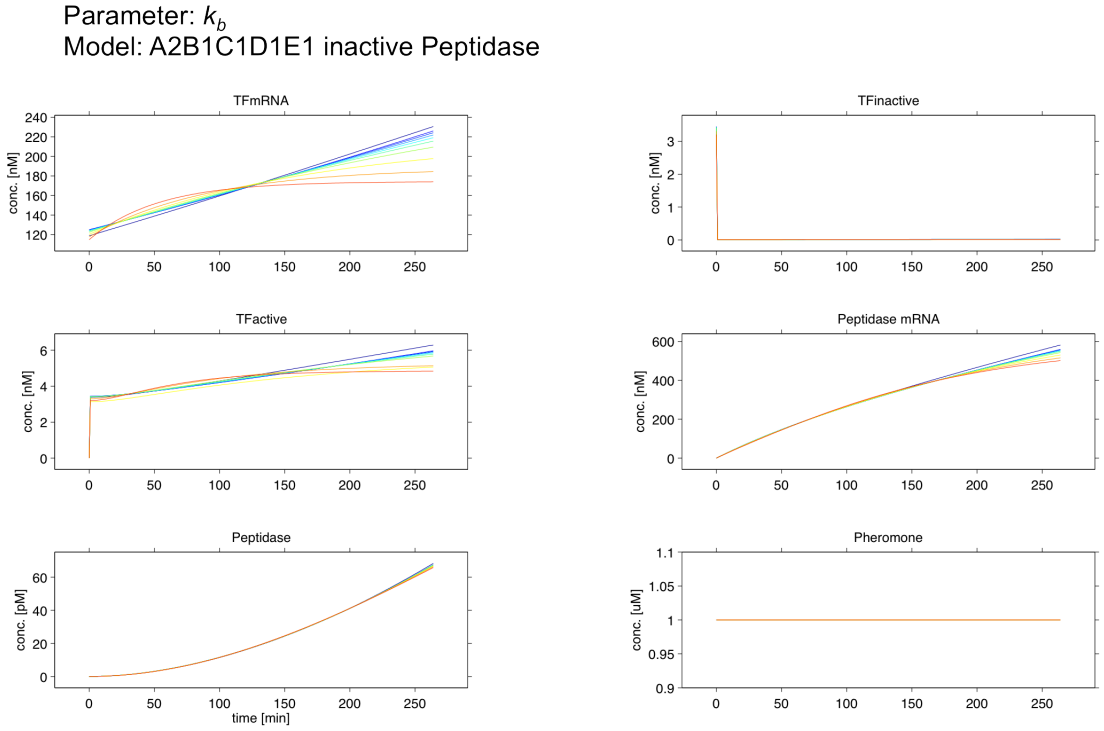


FIGURE A.11: Uncertainty in inactive peptidase model trajectories propagated from uncertainty in k_b.

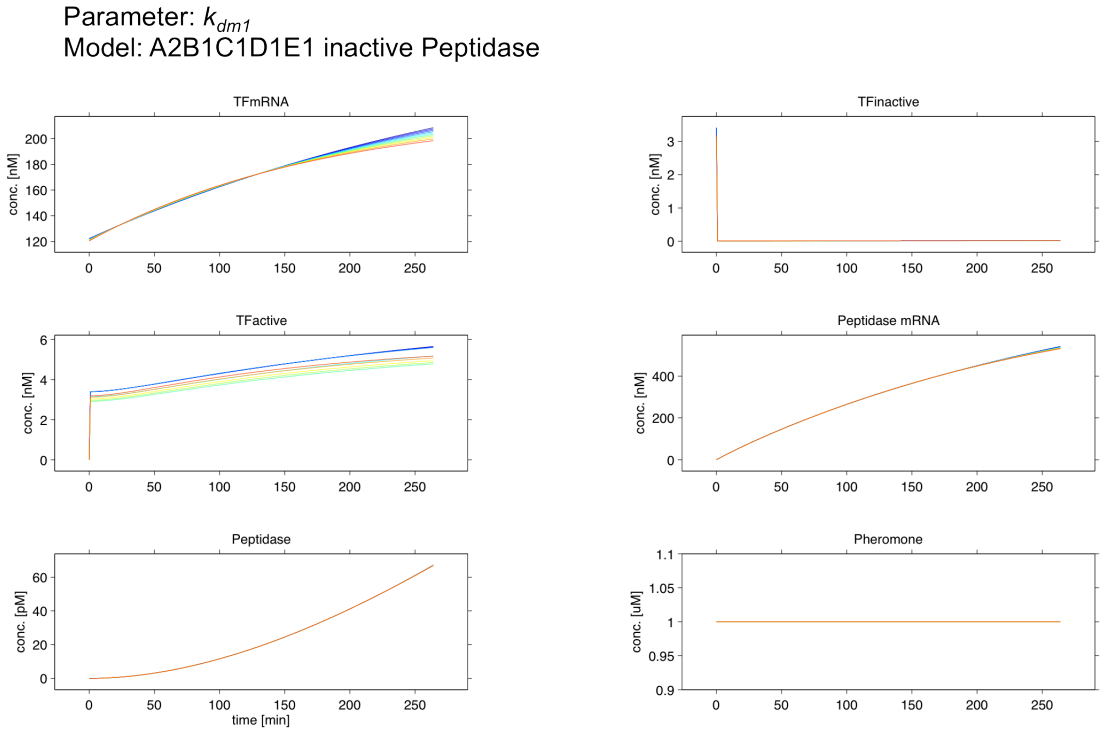


FIGURE A.12: Uncertainty in inactive peptidase model trajectories propagated from uncertainty in $k_{d_{m1}}$.

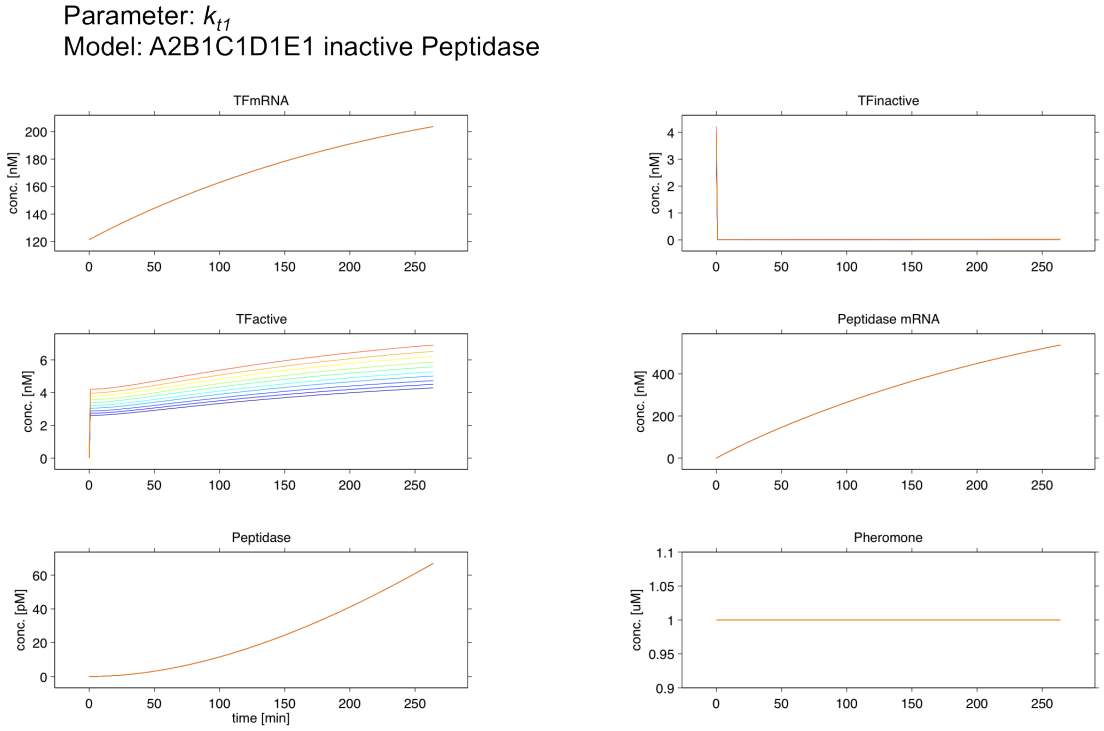


FIGURE A.13: Uncertainty in inactive peptidase model trajectories propagated from uncertainty in k_{t_1}.

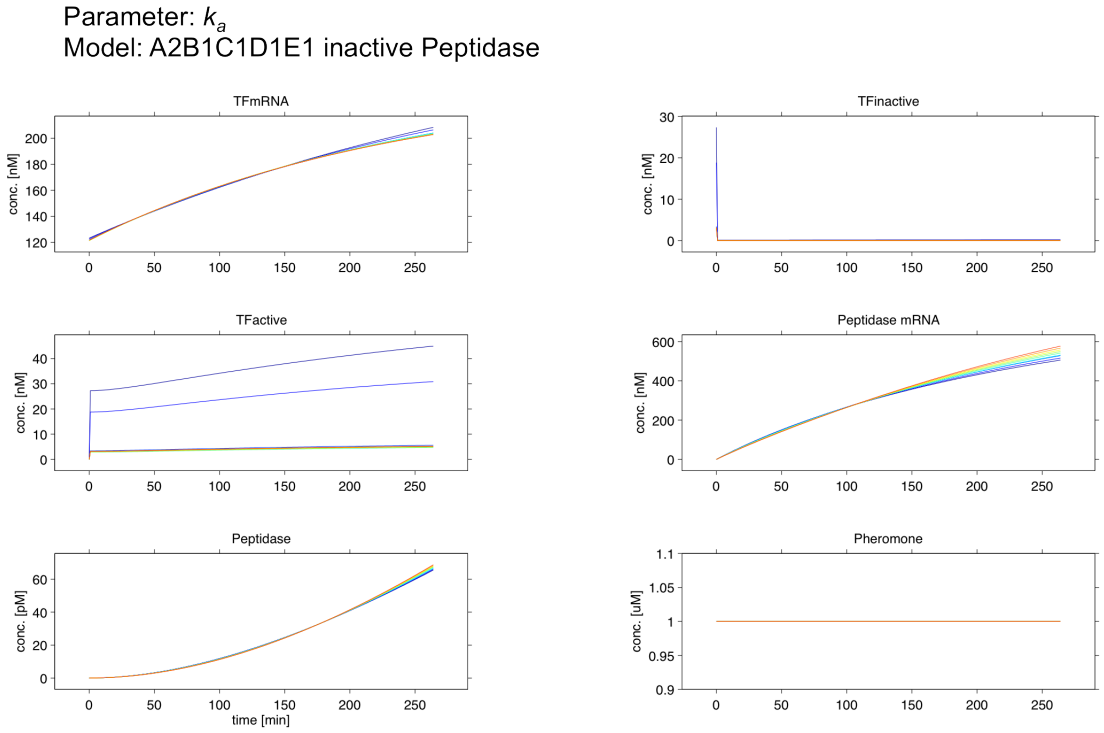


FIGURE A.14: Uncertainty in inactive peptidase model trajectories propagated from uncertainty in k_a.

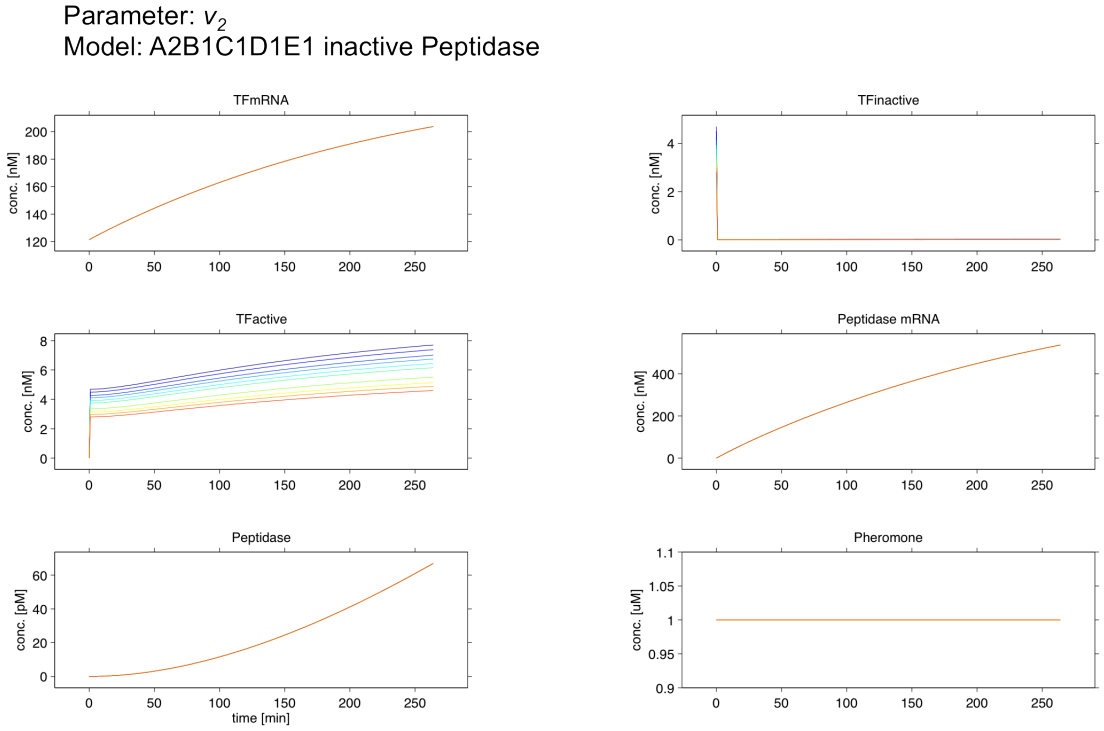


FIGURE A.15: Uncertainty in inactive peptidase model trajectories propagated from uncertainty in v_2.

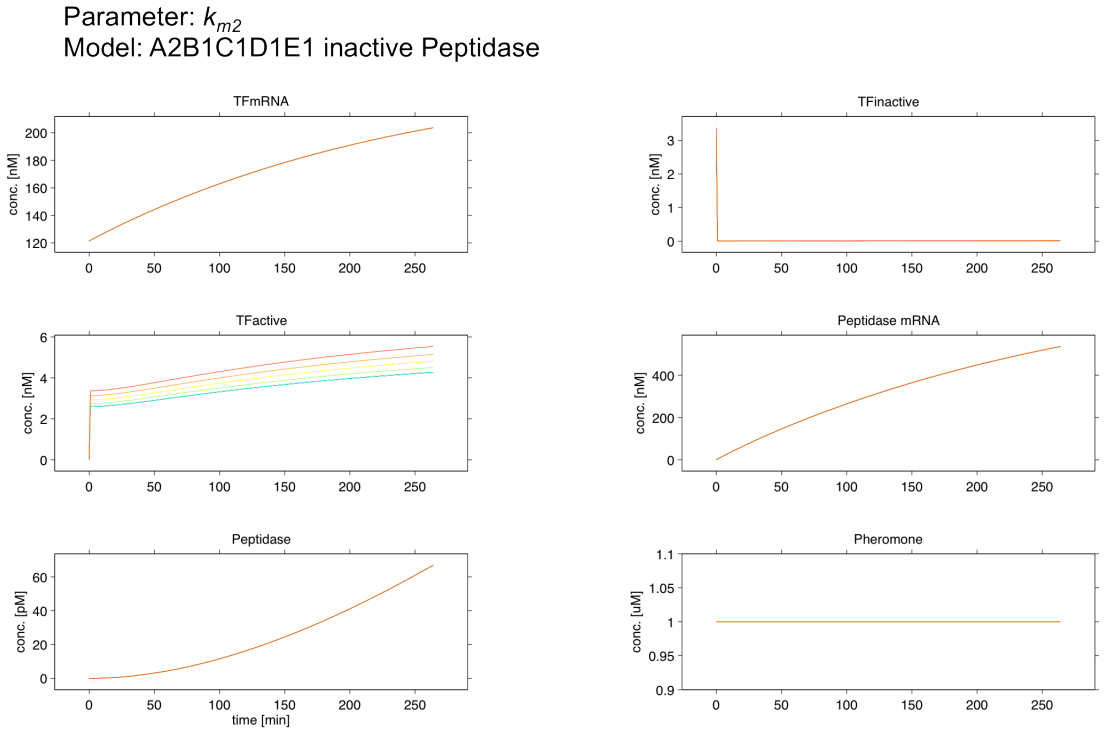


FIGURE A.16: Uncertainty in inactive peptidase model trajectories propagated from uncertainty in k_{m_2}.

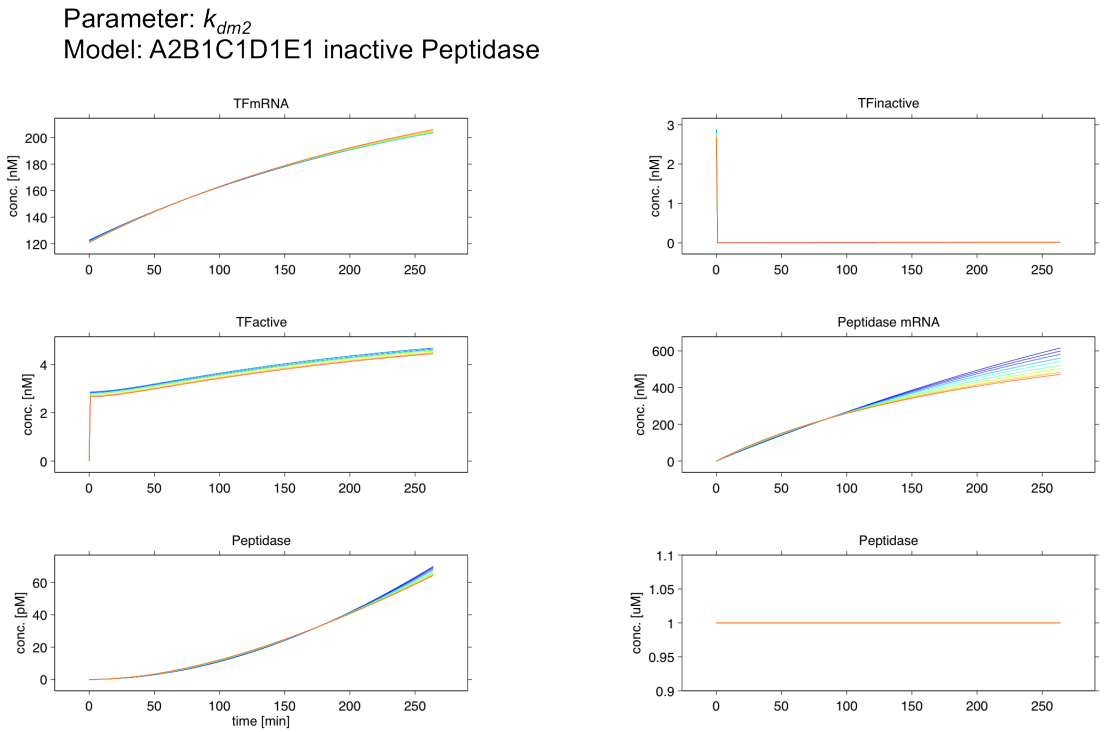


FIGURE A.17: Uncertainty in inactive peptidase model trajectories propagated from uncertainty in $k_{d_{m2}}$.

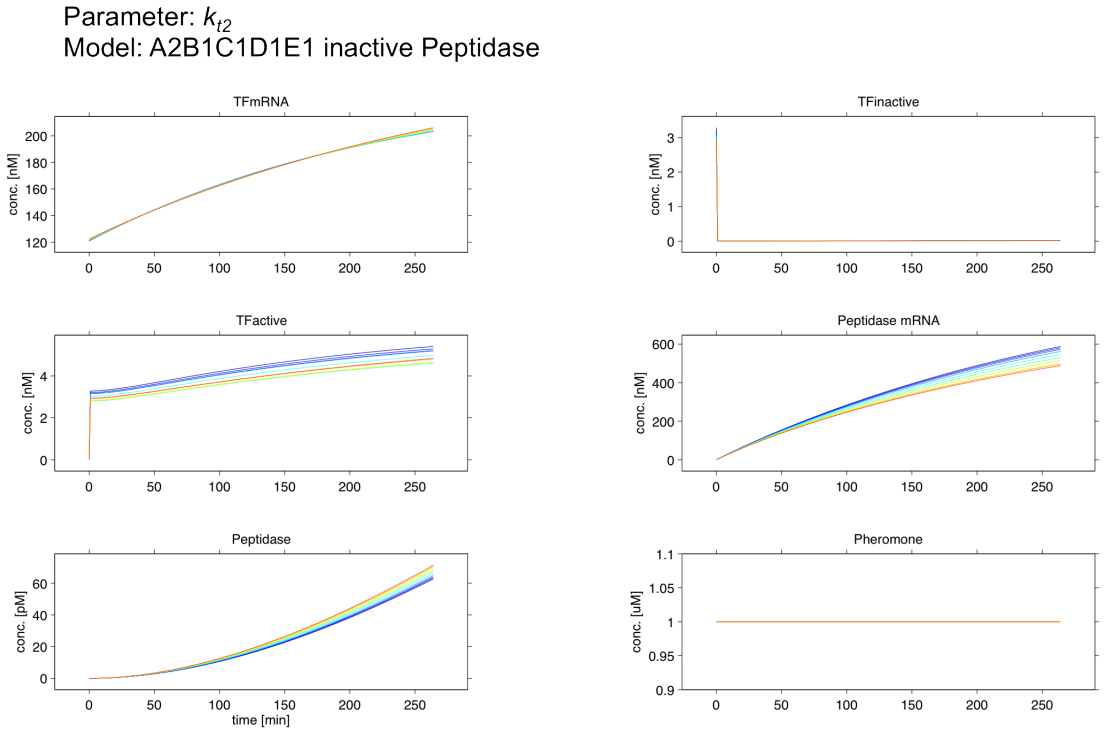


FIGURE A.18: Uncertainty in inactive peptidase model trajectories propagated from uncertainty in k_{t_2}.

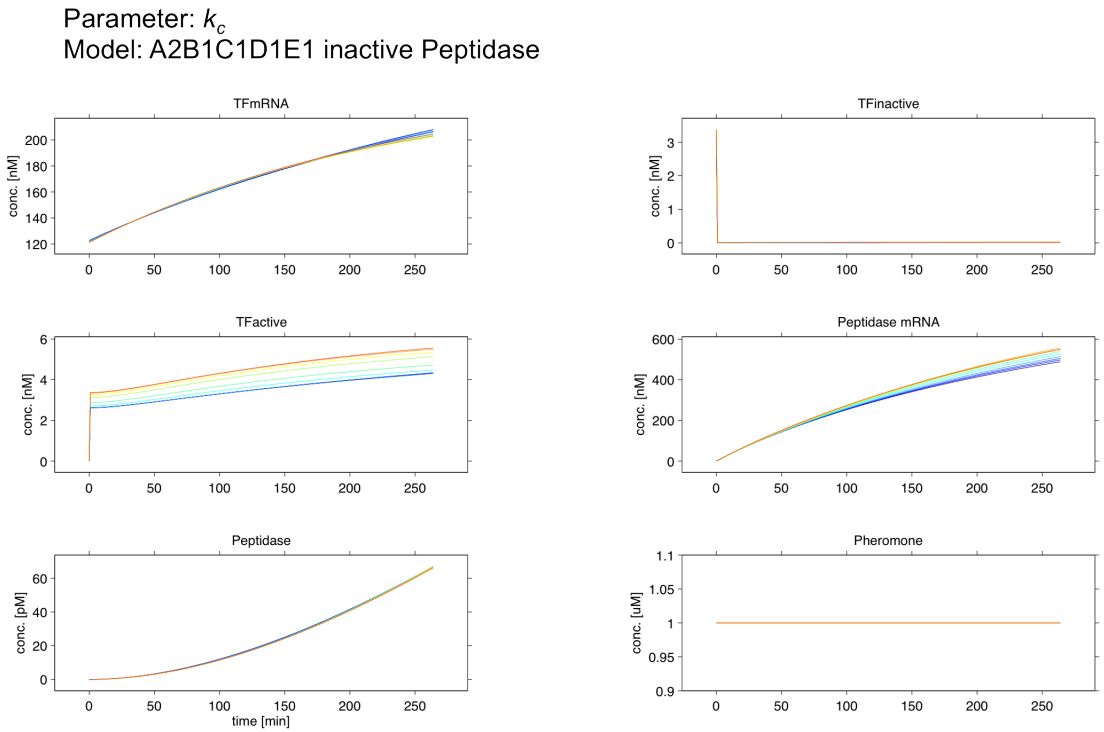


FIGURE A.19: Uncertainty in inactive peptidase model trajectories propagated from uncertainty in k_c.

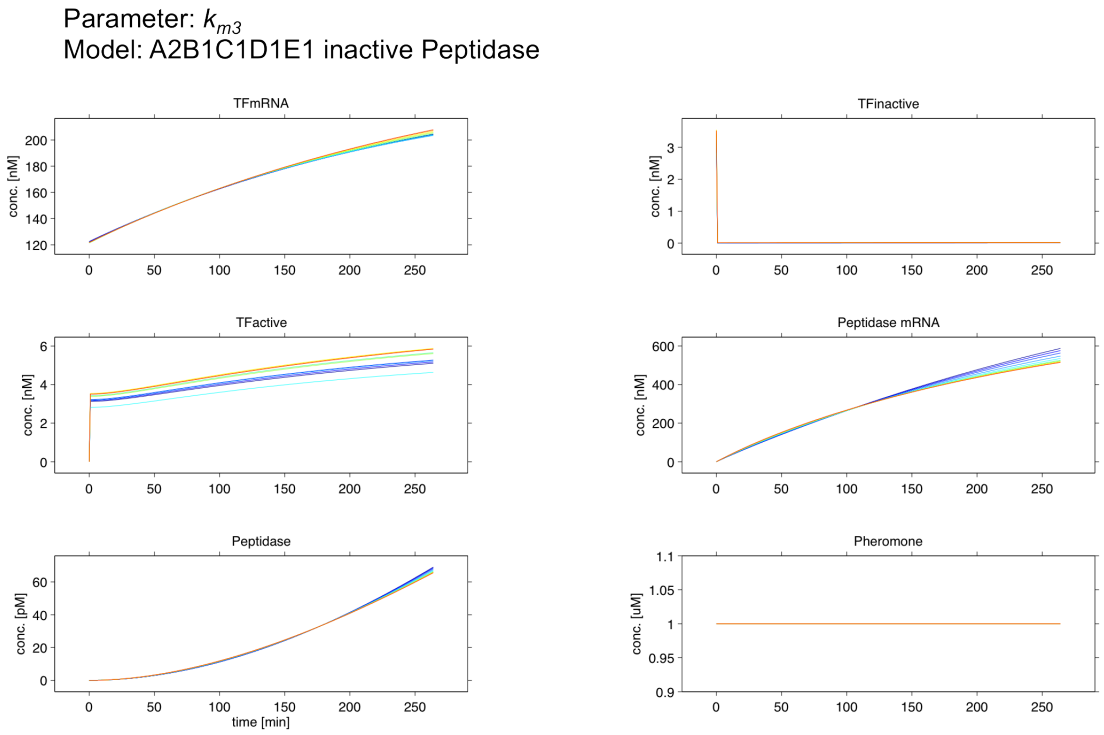


FIGURE A.20: Uncertainty in inactive peptidase model trajectories propagated from uncertainty in k_{m_3}.

Appendix B

Uncertainty in trajectories of TF total

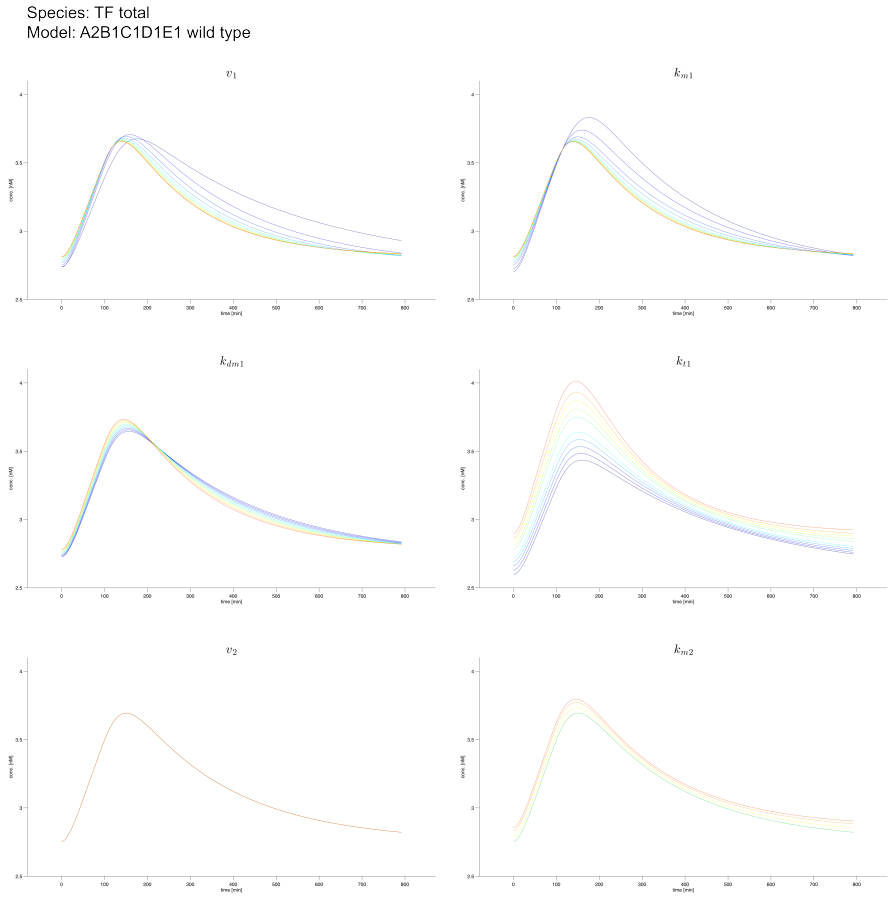


FIGURE B.1: Uncertainty in trajectories of TF total in a wild type simulation.

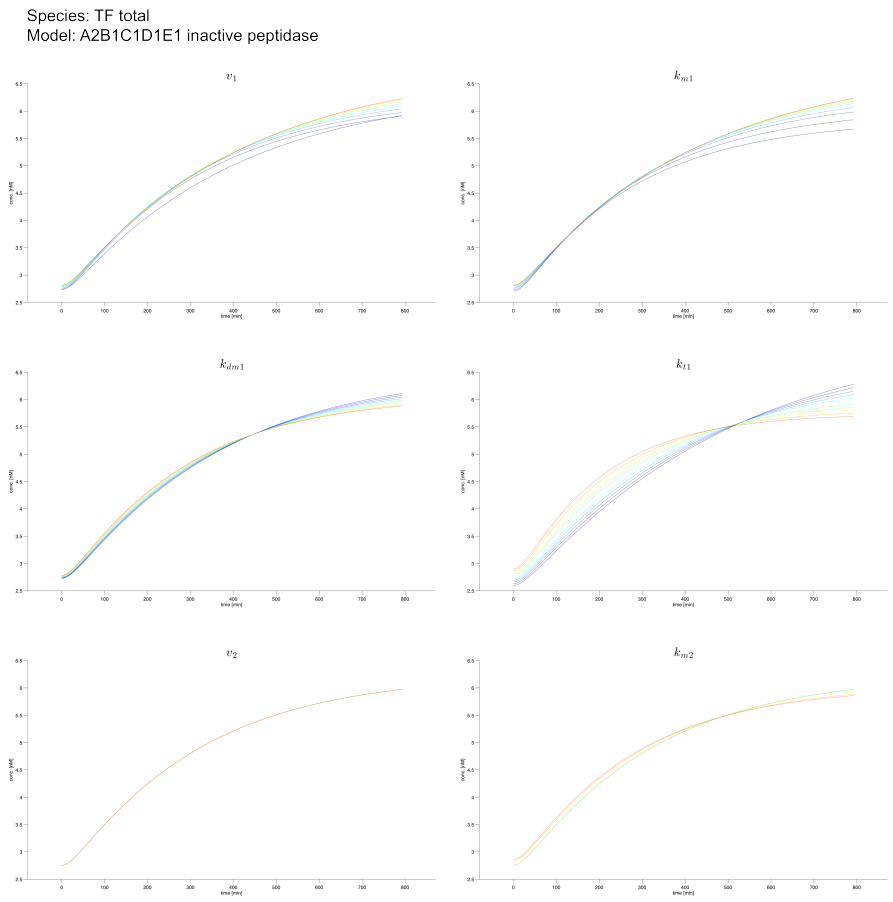


FIGURE B.2: Uncertainty in trajectories of TF total in an inactive peptidase simulation.

Appendix C

PLE of model A2B1C1D1E1 fitted to synthetic data of TF total

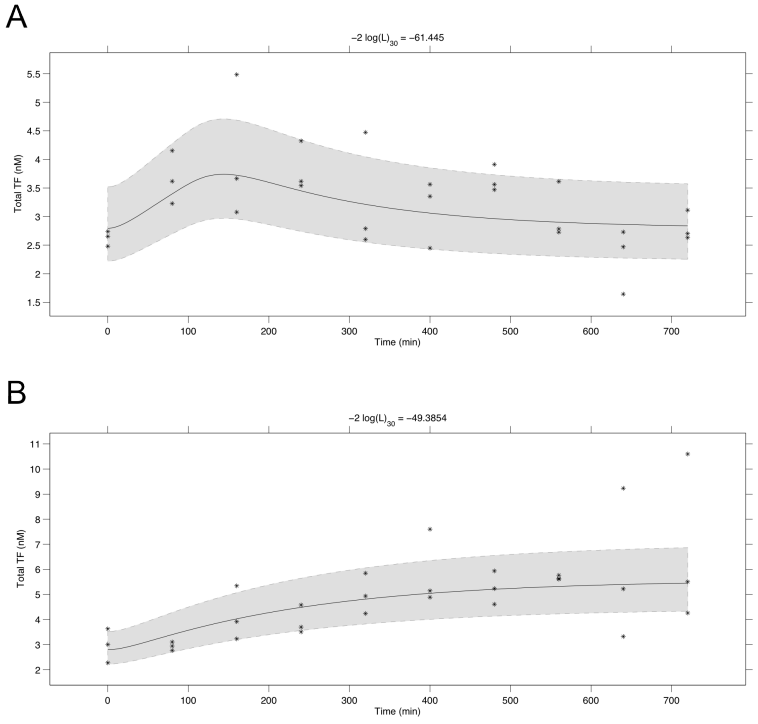


FIGURE C.1: Model A2B1C1D1E1 fit to TF total synthetic data. (A) Wild type condition. (B) Inactive peptidase condition.

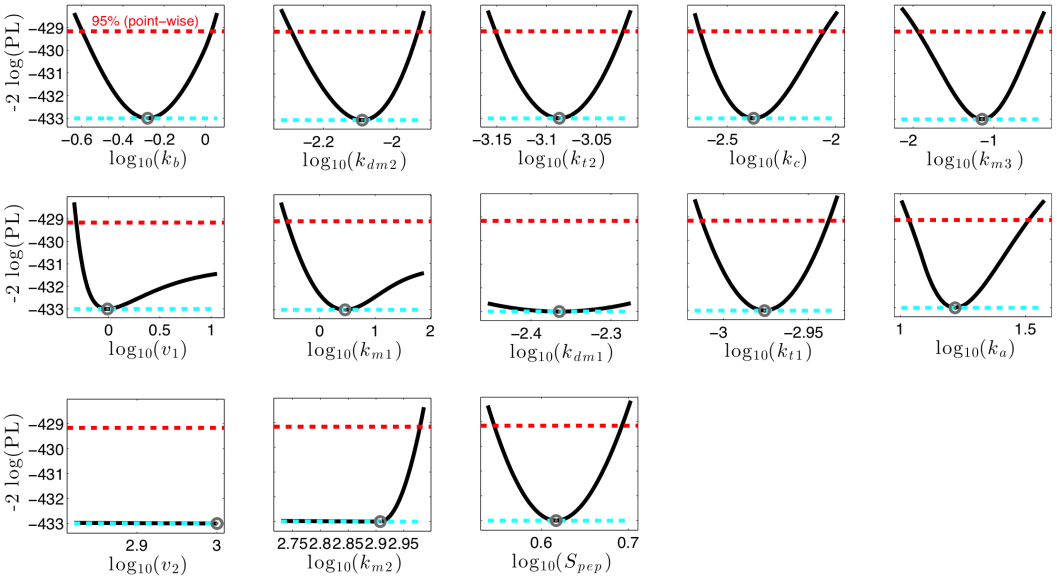


FIGURE C.2: Parameter profile likelihoods for model A2B1C1D1E1 fitted to experimental data including synthetic data of TF total. Estimated optimum values are marked with a circle.

Appendix D

Fitting the reduced model A2 to experimental data

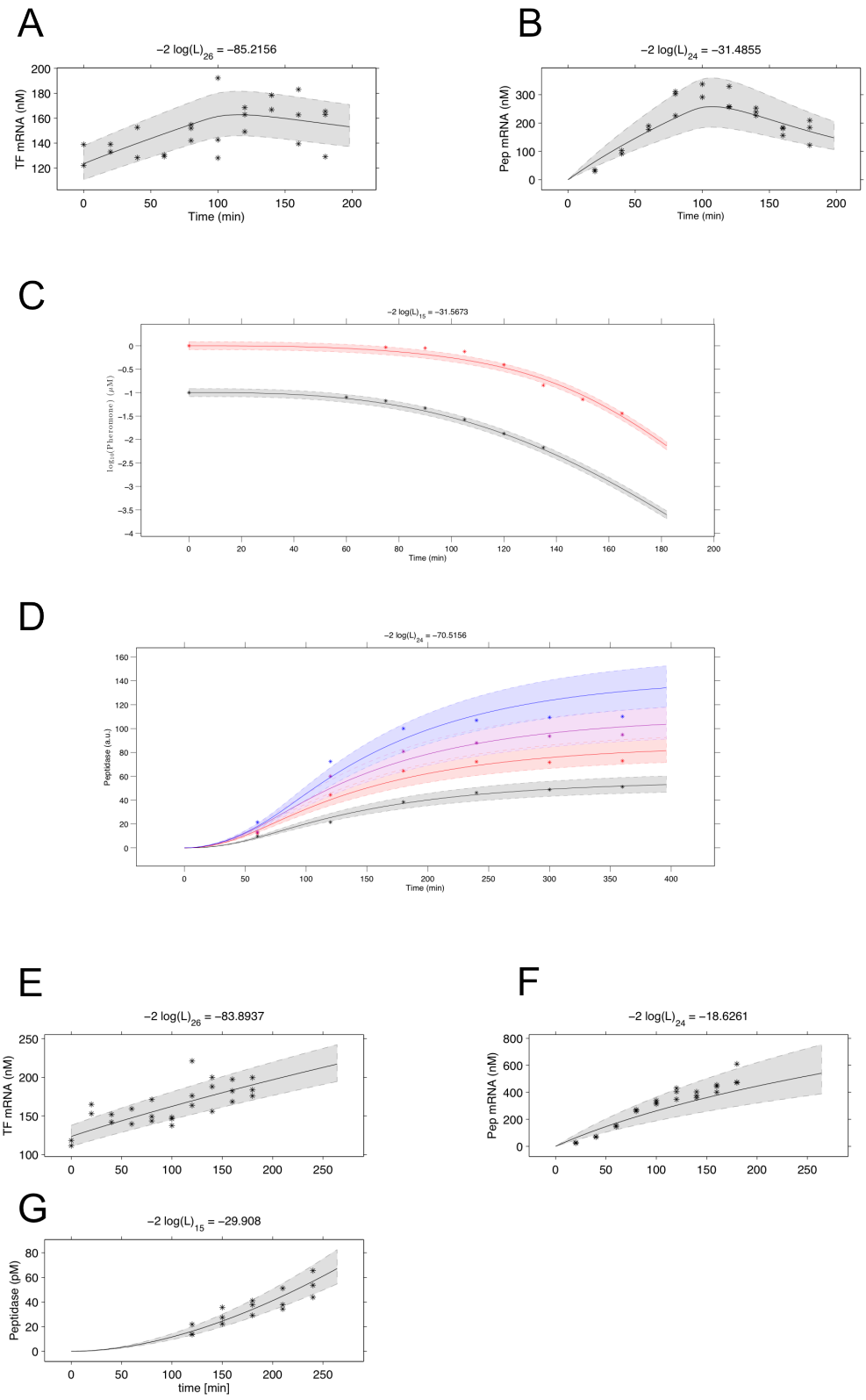


FIGURE D.1: Fitting the reduced model A2 to experimental data. (A-B) Absolute quantification of *ste11* (A) and *sxa2* (B) gene expression for the wild type strain stimulated with 1 µM P-factor pheromone. (C) Absolute quantification of pheromone inactivation for the wild type strain stimulated with 1 µM (black) and 0.1 µM (red) P-factor pheromone. (D) Relative quantification of Sxa2 from Ladds *et al.* (1996). (E-F) Absolute quantification of *ste11* (E) and *sxa2* (F) gene expression for the $Sxa2^{S200A}$ mutant strain stimulated with 1 µM P-factor pheromone. (G) Absolute quantification of $Sxa2^{S200A}$ stimulated with 1 µM P-factor pheromone. Solid traces are the model trajectories fitted to experimental data (dots). Shades are the estimated experimental error of the data.

Appendix E

Image analysis of single cell transcriptional response to pheromone

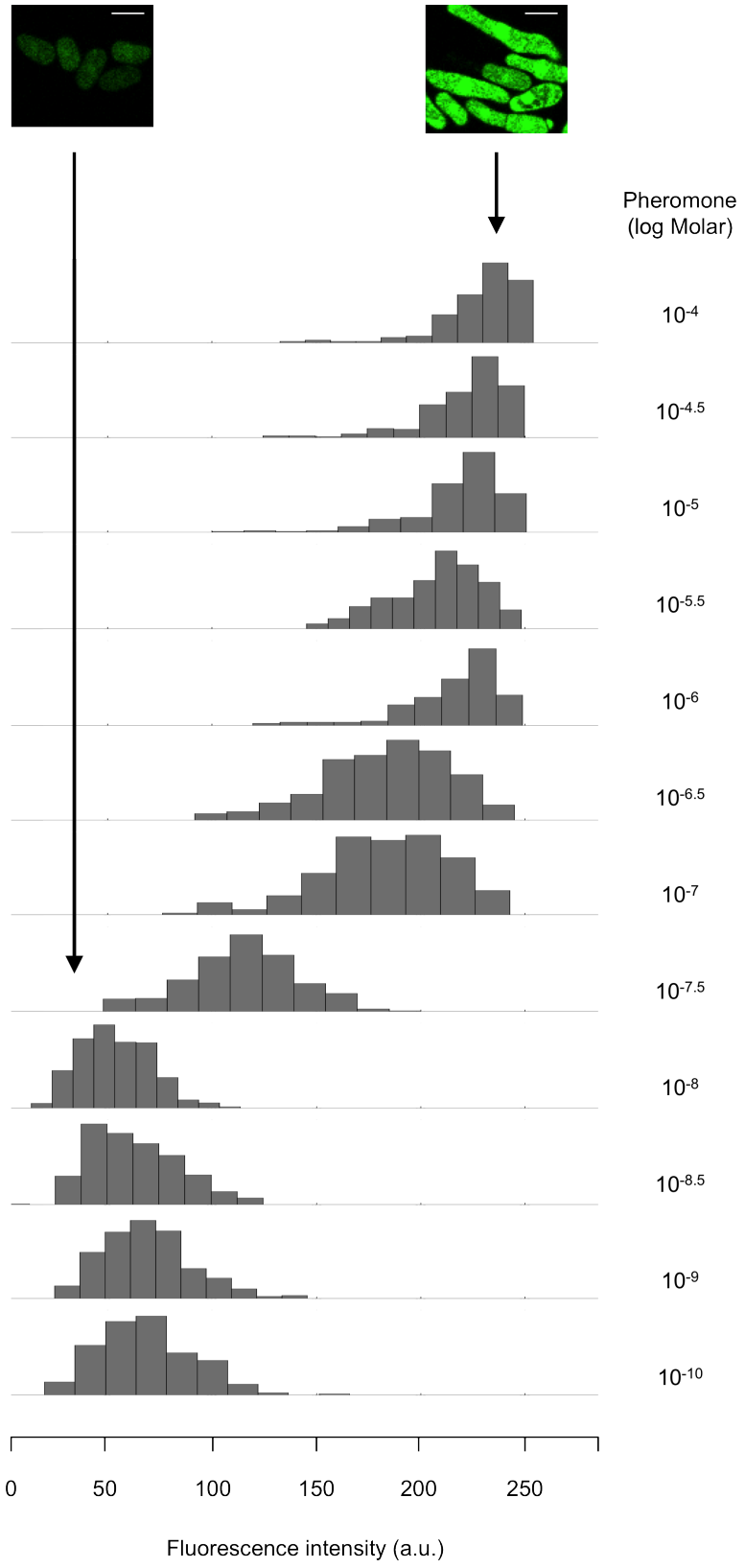


FIGURE E.1: Distributions of single cell reporter expression in response to pheromone. The yeast strain JY1325 described by Smith (2009) harbouring a GFP reporter at the *sxa2* locus, was cultured under standard conditions and stimulated with various concentrations of pheromone for 16 hrs. Histograms show the distribution of total fluorescence intensity measured for individual cells from each dose. Images show representative examples of cells from a low and high response distribution. Image acquisition and quantification is described in detail in the materials and methods section. Scale bars, 10 µm.

Bibliography

Aho K., Derryberry D. and Peterson T. (**2014**). Model selection for ecologists: The worldviews of AIC and BIC. *Ecology*, **95**(3): 631–636.

Ahrén B. (**2009**). Islet G protein-coupled receptors as potential targets for treatment of type 2 diabetes. *Nat. Rev. Drug Discov.*, **8**(5): 369–385.

Akaike H. (**1974**). A new look at the statistical model identification. *IEEE Trans. Autom. Control*, **19**(6): 716–723.

Alberts B., Johnson A., Lewis J., Raff M., Roberts K. and Walter P. (**2008**). *Molecular biology of the cell.* Garland Science, 5th edition.

Alfa C., Fantes P., Hyams J., McLeod M. and Warbrick E. (**1993**). *Experiments with fission yeast: A laboratory manual.* Cold Spring Harbor Laboratory Press.

Alvarez B. and Moreno S. (**2006**). Fission yeast Tor2 promotes cell growth and represses cell differentiation. *J. Cell. Sci.*, **119**(Pt 21): 4475–4485.

Alves R., Vilaprinyo E., Hernádez-Bermejo B. and Sorribas A. (**2008**). Mathematical formalisms based on approximated kinetic representations for modeling genetic and metabolic pathways. *Biotechnol. Genet. Eng. Rev.*, **25**: 1–40.

Anandhakumar J., Fauquenoy S., Materne P., Migeot V. and Hermand D. (**2013**). Regulation of entry into gametogenesis by Ste11: the endless game. *Biochem. Soc. Trans.*, **41**(6): 1673–1678.

Andersen C.L., Jensen J.L. and Ørntoft T.F. (**2004**). Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.*, **64**(15): 5245–5250.

Arikawa E., Sun Y., Wang J., Zhou Q., Ning B., Dial S.L., Guo L. and Yang J. (**2008**). Cross-platform comparison of SYBR Green real-time PCR with TaqMan PCR, microarrays and other gene expression measurement technologies evaluated in the MicroArray Quality Control (MAQC) study. *BMC Genomics*, **9**: 328.

Arvidsson S., Kwasniewski M., Riaño-Pachón D.M. and Mueller-Roeber B. (**2008**). QuantPrime–a flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinformatics*, **9**: 465.

Ay A. and Arnosti D.N. (**2011**). Mathematical modeling of gene expression: a guide for the perplexed biologist. *Crit. Rev. Biochem. Mol.*, **46**(2): 137–151.

Baker S.M., Poskar C.H., Schreiber F. and Junker B.H. (**2015**). A unified framework for estimating parameters of kinetic biological models. *BMC Bioinformatics*, **16**: 104.

Balsa-Canto E., Alonso A.A. and Banga J.R. (**2008**). Computational procedures for optimal experimental design in biological systems. *IET Syst. Biol.*, **2**(4): 163–172.

Balsa-Canto E. and Banga J.R. (**2010**). Computational procedures for model identification. In S. Choi (ed.), *Systems biology for signaling networks.*, Springer. 111–137.

Bazigou E. and Rallis C. (**2007**). Cell signaling and cancer. *Genome Biol.*, **8**(7): 310.

Behar M., Hao N., Dohlman H.G. and Elston T.C. (**2008**). Dose-to-duration encoding and signaling beyond saturation in intracellular signaling networks. *PLoS Comput. Biol.*, **4**(10): e1000197.

Bellman R. and Åström K.J. (**1970**). On structural identifiability. *Math. Biosci.*

Boeke J.D., Lacroute F. and Fink G.R. (**1984**). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.*, **197**: 345–346.

Bornholdt S. (**2005**). Less is more in modeling large genetic networks. *Science*, **310**(5747): 449–451.

Burnham K.P. and Anderson D.R. (**2002**). *Model selection and multimodel inference: A practical information-theoretic approach.* Springer-Verlag, 2nd edition.

Bustin S.A. (**2000**). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.*, **25**(2): 169–193.

Cargnello M. and Roux P.P. (**2011**). Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. *Microbiol. Mol. Biol. Rev.*, **75**(1): 50–83.

Chandran D., Copeland W.B., Sleight S.C. and Sauro H.M. (**2008**). Mathematical modeling and synthetic biology. *Drug Discov. Today Dis. Models*, **5**(4): 299–309.

Chen D., Toone W.M., Mata J., Lyne R., Burns G., Kivinen K., Brazma A., Jones N. and Bähler J. (**2003**). Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell*, **14**(1): 214–229.

Chis O.T., Banga J.R. and Balsa-Canto E. (**2011**). Structural identifiability of systems biology models: a critical comparison of methods. *PLoS ONE*, **6**(11): e27755.

Cobelli C. and DiStefano III J.J. (**1980**). Parameter and structural identifiability concepts and ambiguities: A critical review and analysis. *Am. J. Physiol.*, **239**(1): R7–24.

Collins M.L., Zayati C., Detmer J.J., Daly B., Kolberg J.A., Cha T.A., Irvine B.D., Tucker J. and Urdea M.S. (**1995**). Preparation and characterization of RNA standards for use in quantitative branched DNA hybridization assays. *Anal. Biochem.*, **226**(1): 120–129.

Coudreuse D., van Bakel H., Dewez M., Soutourina J., Parnell T., Vandenhaute J., Cairns B., Werner M. and Hermand D. (**2010**). A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in *S. pombe*. *Curr. Biol.*, **20**(12): 1053–1064.

Coulon A., Chow C.C., Singer R.H. and Larson D.R. (**2013**). Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.*, **14**(8): 572–584.

Croft W., Hill C., McCann E., Bond M., Esparza-Franco M.A., Bennett J., Rand D., Davey J. and Ladds G. (**2013**). A physiologically required G protein-coupled receptor (GPCR)-Regulator of G protein signaling (RGS) interaction that compartmentalizes RGS activity. *J. Biol. Chem.*, **288**(38): 27327–27342.

Croft W.D. (**2012**). *A systems and molecular analysis of G protein-mediated signalling.* Ph.D. thesis, University of Warwick, UK.

Das A., Forfar R., Ladds G. and Davey J. (**2006**). Combined use of two transcriptional reporters improves signalling assays for G protein-coupled receptors in fission yeast. *Yeast*, **23**(12): 889–897.

Davey J. (**1991**). Isolation and quantitation of M-factor, a diffusible mating factor from the fission yeast *Schizosaccharomyces pombe*. *Yeast*, **7**: 357–366.

Davey J. (**1992**). Mating pheromones of the fission yeast *Schizosaccharomyces pombe*: Purification and structural characterization of M-factor and isolation and analysis of two genes encoding the pheromone. *EMBO J.*, **11**(3): 951–960.

Davey J. (**1998**). Fusion of a fission yeast. *Yeast*, **14**(16): 1529–1566.

Davey J., Egel R. and Nielsen O. (**1995**). Pheromone procedures in fission yeast. *Method. Mol. Genet.*, **6**: 247–263.

Davidson E. and Levine M. (**2005**). Gene regulatory networks. *Proc. Natl. Acad. Sci. USA*, **102**: 4935.

Degols G., Shiozaki K. and Russell P. (**1996**). Activation and regulation of the Spc1 stress-activated protein kinase in Schizosaccharomyces pombe. *Mol. Cell. Biol.*, **16**(6): 2870–2877.

Del Vecchio D. (**2011**). The impact of retroactivity on the behavior of biomolecular systems. In *Design and analysis of biomolecular circuits.*, Springer. 161–181.

Del Vecchio D., Ninfa A.J. and Sontag E.D. (**2008**). Modular cell biology: Retroactivity and insulation. *Mol. Syst. Biol.*, **4**.

DeVoti J., Seydoux G., Beach D. and McLeod M. (**1991**). Interaction between *ran1*$^+$ protein kinase and cAMP dependent protein kinase as negative regulators of fission yeast meiosis. *EMBO J.*, **10**(12): 3759–3768.

Didmon M., Davis K., Watson P., Ladds G., Broad P. and Davey J. (**2002**). Identifying regulators of pheromone signalling in the fission yeast *Schizosaccharomyces pombe*. *Curr. Genet.*, **41**(4): 241–253.

Driver R.D. (**1977**). *Ordinary and delay differential equations.* Springer-Verlag.

Egel R. (**1971**). Physiological aspects of conjugation in fission yeast. *Planta*, **98**: 89–96.

Evans N.D., Chapman M.J., Chappell M.J. and Godfrey K.R. (**2002**). Identifiability of uncontrolled nonlinear rational systems. *Automatica*, **38**: 1799–1805.

Evans N.D., Chappell M.J., Chapman M.J. and Godfrey K.R. (**2004**). Structural indistinguishability between uncontrolled (autonomous) nonlinear analytic systems. *Automatica*, **40**(11): 1947–1953.

Falke J.J., Bass R.B., Butler S.L., Chervitz S.A. and Danielson M.A. (**1997**). The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu. Rev. Cell Dev. Biol.*, **13**: 457–512.

Ford J.C., Al-Khodairy F., Fotou E., Sheldrick K.S., Griffiths D.J.F. and Carr A.M. (**1994**). 14-3-3 protein homologs required for the DNA damage checkpoint in fission yeast. *Science*, **265**(5171): 533–535.

Fukui Y., Kaziro Y. and Yamamoto M. (**1986**). Mating pheromone-like diffusible factor released by *Schizosaccharomyces pombe*. *EMBO J.*, **5**(8): 1991–1993.

Gaits F., Degols G., Shiozaki K. and Russell P. (**1998**). Phosphorylation and association with the transcription factor Atf1 regulate localization of Spc1/Sty1 stress-activated kinase in fission yeast. *Genes Dev.*, **12**(10): 1464–1473.

Gibson D.G., Young L., Chuang R.Y., Venter J.C., Hutchison C.A. and Smith H.O. (**2009**). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**(5): 343–345.

Gilman A.G. (**1987**). G proteins: transducers of receptor-generated signals. *Annu. Rev. Biochem.*, **56**: 615–649.

Grandjean T.R.B. (**2013**). *Mathematical modelling of transporter kinetics.* Ph.D. thesis, University of Warwick, UK.

Grimm C., Kohli J., Murray J. and Maundrell K. (**1988**). Genetic engineering of *Schizosaccharomyces pombe*: A system for gene disruption and replacement using the *ura4* gene as a selectable marker. *Mol. Gen. Genet.*, **215**(1): 81–86.

Gurdon J.B., Lemaire P. and Kato K. (**1993**). Community effects and related phenomena in development. *Cell*, **75**(5): 831–834.

Hao N. and O'Shea E.K. (**2012**). Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat. Struct. Mol. Biol.*, **19**(1): 31–39.

Harigaya Y. and Yamamoto M. (**2007**). Molecular mechanisms underlying the mitosis-meiosis decision. *Chromosome Res.*, **15**(5): 523–537.

Hattersley J.G., Pérez-Velázquez J., Chappell M.J., Bearup D., Roper D., Dowson C., Bugg T. and Evans N.D. (**2011**). Indistinguishability and identifiability of kinetic models for the MurC reaction in peptidoglycan biosynthesis. *Comput. Meth. Prog. Bio.*, **104**(2): 70–80.

Hengl S., Kreutz C., Timmer J. and Maiwald T. (**2007**). Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics*, **23**(19): 2612–2618.

Higuchi T., Watanabe Y. and Yamamoto M. (**2002**). Protein kinase A regulates sexual development and gluconeogenesis through phosphorylation of the Zn finger transcriptional activator Rst2p in fission yeast. *Mol. Cell. Biol.*, **22**(1): 1–11.

Hindson B.J., Ness K.D., Masquelier D.A., Belgrader P., Heredia N.J., Makarewicz A.J., Bright I.J., Lucero M.Y., Hiddessen A.L., Legler T.C., Kitano T.K., Hodel M.R., Petersen J.F., Wyatt P.W., Steenblock E.R., Shah P.H., Bousse L.J., Troup C.B., Mellen J.C., K W.D., Erndt N.G., Cauley T.H., Koehler R.T., So A.P., Dube S., Rose K.A., Montesclaros L., Wang S., Stumbo D.P., Hodges S.P., Romine S., Milanovich F.P., White H.E., Regan J.F., Karlin-Neumann G.A., Hindson C.M., Saxonov S. and Colston B.W. (**2011**). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.*, **83**(22): 8604–8610.

Hindson C.M., Chevillet J.R., Briggs H.A., Gallichotte E.N., Ruf I.K., Hindson B.J., Vessella R.L. and Tewari M. (**2013**). Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Methods*, **10**(10): 1003–1005.

Hnasko T.S. and Hnasko R.M. (**2015**). The western blot. In R. Hnasko (ed.), *ELISA*., Springer. 87–96.

Hoffman C.S. (**2005**). Glucose sensing via the protein kinase A pathway in *Schizosaccharomyces pombe*. *Biochem. Soc. Trans.*, **33**(Pt 1): 257–260.

Hughes D.A., Fukui Y. and Yamamoto M. (**1990**). Homologous activators of *ras* in fission and budding yeast. *Nature*, **344**(6264): 355–357.

Imai Y., Miyake S., Hughes D.A. and Yamamoto M. (**1991**). Identification of a GTPase-activating protein homolog in *Schizosaccharomyces pombe*. *Mol. Cell. Biol.*, **11**(6): 3088–3094.

Imai Y. and Yamamoto M. (**1992**). *Schizosaccharomyces pombe sxa1*$^+$ and *sxa2*$^+$ encode putative proteases involved in the mating response. *Mol. Cell. Biol.*, **12**(4): 1827–1834.

Imai Y. and Yamamoto M. (**1994**). The fission yeast mating pheromone P-factor: its molecular structure, gene structure, and ability to induce gene expression and G_1 arrest in the mating partner. *Genes Dev.*, **8**(3): 328–338.

Ingalls B.P. and Sauro H.M. (**2003**). Sensitivity analysis of stoichiometric networks: An extension of metabolic control analysis to non-steady state trajectories. *J. Theor. Biol.*, **222**(1): 23–36.

Ivey F.D. and Hoffman C.S. (**2005**). Direct activation of fission yeast adenylate cyclase by the Gpa2 Galpha of the glucose signaling pathway. *Proc. Natl. Acad. Sci. USA*, **102**(17): 6108–6113.

Jaqaman K. and Danuser G. (**2006**). Linking data to models: Data regression. *Nat. Rev. Mol. Cell Biol.*, **7**(11): 813–819.

Kanoh J., Watanabe Y., Ohsugi M., Iino Y. and Yamamoto M. (**1996**). *Schizosaccharomyces pombe gad7*$^+$ encodes a phosphoprotein with a bZIP domain, which is required for proper G1 arrest and gene expression under nitrogen starvation. *Genes Cells*, **1**(4): 391–408.

Karlebach G. and Shamir R. (**2008**). Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.*, **9**(10): 770–780.

Kawamukai M., Ferguson K., Wigler M. and Young D. (**1991**). Genetic and biochemical analysis of the adenylyl cyclase of Schizosaccharomyces pombe. *Cell Regul.*, **2**(2): 155–164.

Kitamura K., Katayama S., Dhut S., Sato M., Watanabe Y., Yamamoto M. and Toda T. (**2001**). Phosphorylation of Mei2 and Ste11 by Pat1 kinase

inhibits sexual differentiation via ubiquitin proteolysis and 14-3-3 protein in fission yeast. *Dev. Cell*, **1**(3): 389–399.

Kitamura K. and Shimoda C. (**1991**). The *Schizosaccharomyces pombe mam2* gene encodes a putative pheromone receptor which has a significant homology with the *Saccharomyces cerevisiae* Ste2 protein. *EMBO J.*, **10**(12): 3743–3751.

Kitano H., Funahashi A., Matsuoka Y. and Oda K. (**2005**). Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.*, **23**(8): 961–966.

Kjærulff S., Andersen N.R., Borup M.T. and Nielsen O. (**2007**). Cdk phosphorylation of the Ste11 transcription factor constrains differentiation-specific transcription to G_1. *Genes Dev.*, **21**(3): 347–359.

Kjærulff S., Lautrup-Larsen I., Truelsen S., Pedersen M. and Nielsen O. (**2005**). Constitutive activation of the fission yeast pheromone-responsive pathway induces ectopic meiosis and reveals Ste11 as a mitogen-activated protein kinase target. *Mol. Cell. Biol.*, **25**(5): 2045–2059.

Klipp E., Liebermeister W., Wierling C., Kowald A., Lehrach H. and Herwig R. (**2009**). *Systems biology: A textbook.* Wiley-Blackwell.

Kreutz C., Bartolome Rodriguez M.M., Maiwald T., Seidl M., Blum H.E., Mohr L. and Timmer J. (**2007**). An error model for protein quantification. *Bioinformatics*, **23**(20): 2747–2753.

Kreutz C. and Timmer J. (**2009**). Systems biology: experimental design. *FEBS J.*, **276**(4): 923–942.

Kunitomo H., Higuchi T., Iino Y. and Yamamoto M. (**2000**). A zinc-finger protein, Rst2p, regulates transcription of the fission yeast *ste11*$^+$ gene, which encodes a pivotal transcription factor for sexual development. *Mol. Biol. Cell*, **11**(9): 3205–3217.

Ladds G. and Davey J. (**2000**). Sxa2 is a serine carboxypeptidase that degrades extracellular P-factor in the fission yeast *Schizosaccharomyces pombe*. *Mol. Microbiol.*, **36**(2): 377–390.

Ladds G., Rasmussen E.M., Young T., Nielsen O. and Davey J. (**1996**). The *sxa2*-dependent inactivation of the P-factor mating pheromone in the fission yeast *Schizosaccharomyces pombe*. *Mol. Microbiol.*, **20**(1): 35–42.

Ladds G.R. (**1998**). *Adaptation with the fission yeast* Schizosaccharomyces pombe*: the characterisation of the secreted serine carboxypeptidase Sxa2.* Ph.D. thesis, University of Warwick, UK.

Le Novère N. (**2015**). Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.*, **16**(3): 146–158.

Lee M.G. and Nurse P. (**1987**). Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. *Nature*, **327**(6117): 31–35.

Leupold U. (**1950**). Die vererbung von homothallie und heterothallie bei *Schizosaccharomyces pombe*. *C. R. Lab. Carlsberg, Sér. physiol.*, **24**: 381–480.

Li P. and McLeod M. (**1996**). Molecular mimicry in development: identification of *ste11*$^+$ as a substrate and *mei3*$^+$ as a pseudosubstrate inhibitor of *ran1*$^+$ kinase. *Cell*, **87**(5): 869–880.

Li Y., Yi M. and Zou X. (**2013**). Identification of the Molecular Mechanisms for Cell-Fate Selection in Budding Yeast through Mathematical Modeling. *Biophys. J.*, **104**(10): 2282–2294.

Lin A. and Liu Z.g. (**2008**). Cell signaling review series. *Cell Res.*, **18**(3): 327.

Liu H. and Naismith J.H. (**2008**). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.*, **8**: 91.

Ljung L. (**1987**). *System identification: Theory for the user.* Prentice Hall.

Lyne R., Burns G., Mata J., Penkett C.J., Rustici G., Chen D., Langford C., Vetrie D. and Bähler J. (**2003**). Whole-genome microarrays of fission yeast: characteristics, accuracy, reproducibility, and processing of array data. *BMC Genomics*, **4**(1): 27.

MacDonald J.T., Barnes C., Kitney R.I., Freemont P.S. and Stan G.B.V. (**2011**). Computational design approaches and tools for synthetic biology. *Integr. Biol.*, **3**(2): 97–108.

Maeda T., Mochizuki N. and Yamamoto M. (**1990**). Adenylyl cyclase is dispensable for vegetative cell growth in the fission yeast *Schizosaccharomyces pombe*. *Proc. Natl. Acad. Sci. USA*, **87**(20): 7814–7818.

Maeda T., Watanabe Y., Kunitomo H. and Yamamoto M. (**1994**). Cloning of the *pka1* gene encoding the catalytic subunit of the cAMP-dependent protein kinase in *Schizosaccharomyces pombe*. *J. Biol. Chem.*, **269**(13): 9632–9637.

Maiwald T. and Timmer J. (**2008**). Dynamical modeling and multi-experiment fitting with PottersWheel. *Bioinformatics*, **24**(18): 2037–2043.

Marguerat S., Schmidt A., Codlin S., Chen W., Aebersold R. and Bähler J. (**2012**). Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell*, **151**(3): 671–683.

Markham N.R. and Zuker M. (**2008**). UNAFold: Software for nucleic acid folding and hybridization. In J.M. Keith (ed.), *Bioinformatics: Structure, function and applications.*, Humana Press. 3–31.

Mata J. and Bähler J. (**2006**). Global roles of Ste11p, cell type, and pheromone in the control of gene expression during early sexual differentiation in fission yeast. *Proc. Natl. Acad. Sci. USA*, **103**(42): 15517–15522.

Mata J., Lyne R., Burns G. and Bähler J. (**2002**). The transcriptional program of meiosis and sporulation in fission yeast. *Nat. Genet.*, **32**(1): 143–147.

Mata J., Wilbrey A. and Bähler J. (**2007**). Transcriptional regulatory network for sexual differentiation in fission yeast. *Genome Biol.*, **8**(10): R217.

Matsuo T., Otsubo Y., Urano J., Tamanoi F. and Yamamoto M. (**2007**). Loss of the TOR kinase Tor2 mimics nitrogen starvation and activates the sexual development pathway in fission yeast. *Mol. Cell. Biol.*, **27**(8): 3154–3164.

Matsuyama A., Arai R., Yashiroda Y., Shirai A., Kamata A., Sekido S., Kobayashi Y., Hashimoto A., Hamamoto M., Hiraoka Y., Horinouchi S. and Yoshida M. (**2006**). ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. *Nat. Biotechnol.*, **24**(7): 841–847.

McLeod M., Shor B., Caporaso A., Wang W., Chen H. and Hu L. (**2000**). Cpc2, a fission yeast homologue of mammalian RACK1 protein, interacts with Ran1 (Pat1) kinase To regulate cell cycle progression and meiotic development. *Mol. Cell. Biol.*, **20**(11): 4016–4027.

Millar J.B.A., Buck V. and Wilkinson M.G. (**1995**). Pyp1 and Pyp2 PTPases dephosphorylate an osmosensing MAP kinase controlling cell size at division in fission yeast. *Genes Dev.*, **9**(17): 2117–2130.

Mochizuki N. and Yamamoto M. (**1992**). Reduction in the intracellular cAMP level triggers initiation of sexual development in fission yeast. *Mol. Gen. Genet.*, **233**(1-2): 17–24.

Morigasaki S., Ikner A., Tatebe H. and Shiozaki K. (**2013**). Response regulator-mediated MAPKKK heteromer promotes stress signaling to the Spc1 MAPK in fission yeast. *Mol. Biol. Cell*, **24**(7): 1083–1092.

Motulsky H.J. and Christopoulos A. (**2004**). *Fitting models to biological data using linear and nonlinear regression: A practical guide to curve fitting.* Oxford University Press.

Neale M.C. and Miller M.B. (**1997**). The use of likelihood-based confidence intervals in genetic models. *Behav. Genet.*, **27**(2): 113–120.

Neiman A.M., Stevenson B.J., Xu H.P., Sprague G.F., Herskowitz I., Wigler M. and Marcus S. (**1993**). Functional homology of protein kinases required for sexual differentiation in *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* suggests a conserved signal transduction module in eukaryotic organisms. *Mol. Biol. Cell*, **4**(1): 107–120.

Nielsen H., Engelbrecht J., Brunak S. and von Heijne G. (**1997**). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.*, **10**(1): 1–6.

Obara T., Nakafuku M., Yamamoto M. and Kaziro Y. (**1991**). Isolation and characterization of a gene encoding a G-protein α subunit from *Schizosaccharomyces pombe*: Involvement in mating and sporulation pathways. *Proc. Natl. Acad. Sci. USA*, **88**(13): 5877–5881.

Okazaki K., Okazaki N., Kume K., Jinno S., Tanaka K. and Okayama H. (**1990**). High-frequency transformation method and library transducing vectors for cloning mammalian cDNAs by *trans*-complementation of *Schizosaccharomyces pombe*. *Nucleic Acids Res.*, **18**(22): 6485–6489.

Oowatari Y., Jeong H., Tanae K., Nakagawa T. and Kawamukai M. (**2011**). Regulation and role of an RNA-binding protein Msa2 in controlling the sexual differentiation of fission yeast. *Curr. Genet.*, **57**(3): 191–200.

Otsubo Y. and Yamamoto M. (**2012**). Signaling pathways for fission yeast sexual differentiation at a glance. *J. Cell. Sci.*, **125**(Pt 12): 2789–2793.

Owen A.B. (**1992**). A central limit theorem for Latin hypercube sampling. *J. Roy. Stat. Soc. B Met.*, **54**(2): 541–551.

Pancaldi V., Saraç Ö.S., Rallis C., McLean J.R., Převorovský M., Gould K., Beyer A. and Bähler J. (**2012**). Predicting the fission yeast protein interaction network. *G3: Genes, Genomes, Genetics*, **2**(4): 453–467.

Papadaki P., Pizon V., Onken B. and Chang E.C. (**2002**). Two ras pathways in fission yeast are differentially regulated by two ras guanine nucleotide exchange factors. *Mol. Cell. Biol.*, **22**(13): 4598–4606.

Pfaffl M.W. (**2001**). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, **29**(9): e45.

Pfaffl M.W., Tichopad A., Prgomet C. and Neuvians T.P. (**2004**). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnol. Lett.*, **26**(6): 509–515.

Phatnani H.P. and Greenleaf A.L. (**2006**). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.*, **20**(21): 2922–2936.

Pohjanpalo H. (**1978**). System identifiability based on the power series expansion of the solution. *Math. Biosci.*, **41**: 21–33.

Purves D., Augustine G.J., Fitzpatrick D., Hall W.C., Lamantia A.S., McNamara J.O. and Williams S.M. (eds.) (**2004**). *Neuroscience.* Sunderland: Sinauer Associates, Inc., 3rd edition.

Purvis J.E. and Lahav G. (**2013**). Encoding and Decoding Cellular Information through Signaling Dynamics. *Cell*, **152**(5): 945–956.

Qin J., Kang W., Leung B. and McLeod M. (**2003**). Ste11p, a high-mobility-group box DNA-binding protein, undergoes pheromone- and nutrient-regulated nuclear-cytoplasmic shuttling. *Mol. Cell. Biol.*, **23**(9): 3253–3264.

Quaiser T. and Mönnigmann M. (**2009**). Systematic identifiability testing for unambiguous mechanistic modeling–application to JAK-STAT, MAP kinase, and NF-κB signaling pathway models. *BMC Syst. Biol.*, **3**: 50.

Quellhorst G. and Rulli S. (**2008**). A systematic guideline for developing the best real-time PCR primers. What we have learned from designing assays for more than 14,000 genes. SABiosciences.

Radonić A., Thulke S., Mackay I.M., Landt O., Siegert W. and Nitsche A. (**2004**). Guideline to reference gene selection for quantitative real-time PCR. *Biochem. Biophys. Res. Commun.*, **313**(4): 856–862.

Raue A., Karlsson J., Saccomani M.P., Jirstrand M. and Timmer J. (**2014**). Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics*, **30**(10): 1440–1448.

Raue A., Kreutz C., Maiwald T., Bachmann J., Schilling M., Klingmüller U. and Timmer J. (**2009**). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, **25**(15): 1923–1929.

Raue A., Kreutz C., Maiwald T., Klingmüller U. and Timmer J. (**2011**). Addressing parameter identifiability by model-based experimentation. *IET Syst. Biol.*, **5**(2): 120.

Raue A., Schilling M., Bachmann J., Matteson A., Schelke M., Kaschek D., Hug S., Kreutz C., Harms B.D., Theis F.J., Klingmüller U. and Timmer J. (**2013**). Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE*, **8**(9): e74335.

Raue A., Steiert B., Schelker M., Kreutz C., Maiwald T., Hass H., Vanlier J., Tönsing C., Adlung L., Engesser R., Mader W., Heinemann T., Hasenauer J., Schilling M., Höfer T., Klipp E., Theis F.,

Klingmüller U., Schöberl B. and Timmer J. (**2015**). Data2Dynamics: A modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics*: In press.

Raue A. and Timmer J. (**2011**). Model identification by utilizing likelihood-based methods. In M.P.H. Stumpf, D.J. Balding and M. Girolami (eds.), *Handbook of statistical systems biology.*, Wiley. 395–416.

Reiter W., Watt S., Dawson K., Lawrence C.L., Bähler J., Jones N. and Wilkinson C.R.M. (**2008**). Fission yeast MAP kinase Sty1 is recruited to stress-induced genes. *J. Biol. Chem.*, **283**(15): 9945–9956.

Roper R.T., Saccomani M.P. and Vicini P. (**2010**). Cellular signaling identifiability analysis: A case study. *J. Theor. Biol.*, **264**(2): 528–537.

Rozen S. and Skaletsky H. (**1999**). Primer3 on the www for general users and for biologist programmers. In S. Misener and S.A. Krawetz (eds.), *Bioinformatics methods and protocols*, Humana Press. 365–386.

Rustici G., Mata J., Kivinen K., Lió P., Penkett C.J., Burns G., Hayles J., Brazma A., Nurse P. and Bähler J. (**2004**). Periodic gene expression program of the fission yeast cell cycle. *Nature Genet.*, **36**(8): 809–817.

Sambrook J. and Russell D.W. (**2001**). *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, 3rd edition.

Samejima I., Mackie S. and Fantes P.A. (**1997**). Multiple modes of activation of the stress-responsive MAP kinase pathway in fission yeast. *EMBO J.*, **16**(20): 6162–6170.

Samejima I., Mackie S., Warbrick E., Weisman R. and Fantes P.A. (**1998**). The fission yeast mitotic regulator *win1*$^+$ encodes an MAP kinase kinase kinase that phosphorylates and activates Wis1 MAP kinase kinase in response to high osmolarity. *Mol. Biol. Cell*, **9**(8): 2325–2335.

Sansó M., Vargas-Pérez I., García P., Ayté J. and Hidalgo E. (**2011**). Nuclear roles and regulation of chromatin structure by the stress-dependent MAP kinase Sty1 of Schizosaccharomyces pombe. *Mol. Microbiol.*, **82**(3): 542–554.

Schenkendorf R., Kremling A. and Mangold M. (**2009**). Optimal experimental design with the sigma point method. *IET Syst. Biol.*, **3**(1): 10–23.

Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., Preibisch S., Rueden C., Saalfeld S., Schmid B., Tinevez J.Y., White D.J., Hartenstein V., Eliceiri K., Tomancak P. and Cardona A. (**2012**). Fiji: an open-source platform for biological-image analysis. *Nat. methods*, **9**(7): 676–682.

Shieh J.C., Wilkinson M.G., Buck V., Morgan B.A., Makino K. and Millar J.B.A. (**1997**). The Mcs4 response regulator coordinately controls the stress-activated Wak1-Wis1-Sty1 MAP kinase pathway and fission yeast cell cycle. *Genes Dev.*, **11**(8): 1008–1022.

Shieh J.C., Wilkinson M.G. and Millar J.B.A. (**1998**). The Win1 mitotic regulator is a component of the fission yeast stress-activated Sty1 MAPK pathway. *Mol. Biol. Cell*, **9**(2): 311–322.

Shiozaki K. and Russell P. (**1995**). Cell-cycle control linked to extracellular environment by MAP kinase pathway in fission yeast. *Nature*, **378**(6558): 739–743.

Shiozaki K. and Russell P. (**1996**). Conjugation, meiosis, and the osmotic stress response are regulated by Spc1 kinase through Atf1 transcription factor in fission yeast. *Genes Dev.*, **10**(18): 2276–2288.

Shiozaki K., Shiozaki M. and Russell P. (**1997**). Mcs4 mitotic catastrophe suppressor regulates the fission yeast cell cycle through the Wik1-Wis1-Spc1 kinase cascade. *Mol. Biol. Cell*, **8**(3): 409–419.

Smith B., Hill C., Godfrey E.L., Rand D., van den Berg H., Thornton S., Hodgkin M., Davey J. and Ladds G. (**2009**). Dual positive and negative regulation of GPCR signaling by GTP hydrolysis. *Cell. Signal.*, **21**(7): 1151–1160.

Smith B.C. (**2009**). *A systems biology analysis of G protein-mediated signalling in eukaryotic cells.* Ph.D. thesis, University of Warwick, UK.

Steiert B., Raue A., Timmer J. and Kreutz C. (**2012**). Experimental design for parameter estimation of gene regulatory networks. *PLoS ONE*, **7**(7): e40052.

Stern B. and Nurse P. (**1996**). A quantitative model for the cdc2 control of S phase and mitosis in fission yeast. *Trends Genet.*, **12**(9): 345–350.

Stern B. and Nurse P. (**1998**). Cyclin B proteolysis and the cyclin-dependent kinase inhibitor rum1p are required for pheromone-induced G1 arrest in fission yeast. *Mol. Biol. Cell*, **9**(6): 1309–1321.

Sugimoto A., Iino Y., Maeda T., Watanabe Y. and Yamamoto M. (**1991**). *Schizosaccharomyces pombe ste11*$^+$ encodes a transcription factor with an HMG motif that is a critical regulator of sexual development. *Genes Dev.*, **5**(11): 1990–1999.

Sukegawa Y., Yamashita A. and Yamamoto M. (**2011**). The Fission Yeast Stress-Responsive MAPK Pathway Promotes Meiosis via the Phosphorylation of Pol II CTD in Response to Environmental and Feedback Cues. *PLoS Genet.*, **7**(12): e1002387.

Tamanoi F. (**2011**). Ras signaling in yeast. *Genes Cancer*, **2**(3): 210–215.

Tanaka K., Davey J., Imai Y. and Yamamoto M. (**1993**). *Schizosaccharomyces pombe map3*$^+$ encodes the putative M-factor receptor. *Mol. Cell. Biol.*, **13**(1): 80–88.

Tesmer J.J.G., Sunahara R.K., Gilman A.G. and Sprang S.R. (**1997**). Crystal structure of the catalytic domains of adenylyl cyclase in a complex with $G_{s\alpha}$·GTPγS. *Science*, **278**(5345): 1907–1916.

Tsukahara K., Yamamoto H. and Okayama H. (**1998**). An RNA binding protein negatively controlling differentiation in fission yeast. *Mol. Cell. Biol.*, **18**(8): 4488–4498.

Untergasser A., Nijveen H., Rao X., Bisseling T., Geurts R. and Leunissen J.A.M. (**2007**). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.*, **35**(suppl. 2): W71–W74.

Uritani M., Hidaka H., Hotta Y., Ueno M., Ushimaru T. and Toda T. (**2006**). Fission yeast Tor2 links nitrogen signals to cell proliferation and acts downstream of the Rheb GTPase. *Genes Cells*, **11**(12): 1367–1379.

Valbuena N. and Moreno S. (**2010**). TOR and PKA pathways synergize at the level of the Ste11 transcription factor to prevent mating and meiosis in fission yeast. *PLoS ONE*, **5**(7): e11514.

van den Berg H. (**2011**). *Mathematical models of biological systems.* Oxford University Press.

VanGuilder H.D., Vrana K.E. and Freeman W.M. (**2008**). Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques*, **44**: 619–626.

Watanabe Y. and Yamamoto M. (**1996**). *Schizosaccharomyces pombe pcr1*$^+$ encodes a CREB/ATF protein involved in regulation of gene expression for sexual development. *Mol. Cell. Biol.*, **16**(2): 704–711.

Weisman R. and Choder M. (**2001**). The fission yeast TOR homolog, *tor1*$^{+}$, is required for the response to starvation and other stresses via a conserved serine. *J. Biol. Chem.*, **276**(10): 7027–7032.

Weisman R., Roitburg I., Schonbrun M., Harari R. and Kupiec M. (**2007**). Opposite effects of *tor1* and *tor2* on nitrogen starvation responses in fission yeast. *Genetics*, **175**(3): 1153–1162.

Weston C., Bond M., Croft W. and Ladds G. (**2013**). The coordination of cell growth during fission yeast mating requires Ras1-GTP hydrolysis. *PLoS ONE*, **8**(10): e77487.

Wilkinson M.G., Samuels M., Takeda T., Toone W.M., Shieh J.C., Toda T., Millar J.B.A. and Jones N. (**1996**). The Atf1 transcription factor is a target for the Sty1 stress-activated MAP kinase pathway in fission yeast. *Genes Dev.*, **10**(18): 2289–2301.

Wittinghofer A. and Vetter I.R. (**2011**). Structure-function relationships of the G domain, a canonical switch motif. *Annu. Rev. Biochem.*, **80**: 943–971.

Wood V., Harris M.A., McDowall M.D., Rutherford K., Vaughan B.W., Staines D.M., Aslett M., Lock A., Bähler J., Kersey P.J. and Oliver S.G. (**2012**). PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res.*, **40**(Database issue): D695–D699.

Xu H.P., White M., Marcus S. and Wigler M. (**1994**). Concerted action of RAS and G proteins in the sexual response pathways of *Schizosaccharomyces pombe*. *Mol. Cell. Biol.*, **14**(1): 50–58.

Xue-Franzén Y., Kjærulff S., Holmberg C., Wright A. and Nielsen O. (**2006**). Genomewide identification of pheromone-targeted transcription in fission yeast. *BMC Genomics*, **7**: 303.

Yi T.M., Kitano H. and Simon M.I. (**2003**). A quantitative characterization of the yeast heterotrimeric G protein cycle. *Proc. Natl. Acad. Sci. USA*, **100**(19): 10764–10769.

Zheng Y. and Sriram G. (**2010**). Mathematical modeling: bridging the gap between concept and realization in synthetic biology. *J. Biomed. Biotechnol.*, **2010**: 541609.

Zi Z. (**2011**). Sensitivity analysis approaches applied to systems biology models. *IET Syst. Biol.*, **5**(6): 336–336.